DOCUMENT RESUME

ED 152 443 95 RC 010 302

TITLE	A Handbook on Evaluation for Title IV Indian Education Act Projects. Reading Supplement: Needs Assessment; Financial Evaluation; The Use of Tests in Program Evaluation; Curriculum Evaluation; Personnel Evaluation and In-Service Training.
INSTITUTION	CPI Associates, Inc., Washington, D.C.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Office of Indian Education.
PUB DATE	76
CONTRACT	OEC-73-7058
NOTE	131p.; For related document, see RC 010 301
EDRS PRICE	MF-$0.83 HC-$7.35 Plus Postage.
DESCRIPTORS	*American Indians; *Cost Effectiveness; Criterion Referenced Tests; *Curriculum Evaluation; Definitions; Federal Legislation; *Guides; *Needs Assessment; Norm Referenced Tests; Personnel Evaluation; *Program Evaluation; School Districts; Tests
IDENTIFIERS	*Indian Education Act 1972

ABSTRACT

A reading supplement to provide background material for "A Handbook on Evaluation for Title IV Indian Education Projects", this document describes and defines the following: (1) needs assessment (definition and purpose; relation of needs to Indian Education Act; approaches; surveying--population sampling, kinds of questions, etc.; group meeting approach; Delphi Technique; panel review; records and archival data; recent trends in needs assessment; and analysis and use); (2) financial evaluation (financial information; financial information and the Indian Education Act; identifying financial information needs--review of project plan, objectives, selected cost accounting, evaluation procedures, etc.; selecting data to be collected--levels of information requirements, availability of data, etc.; and collecting and processing data); (3) the use of tests in program evaluation (criteria; specialized test vocabulary; norm-referenced vs criterion referenced tests; major steps involved in the use of tests--select, review, modify, etc.); (4) curriculum evaluation (basic conceptualizations; special projects; models; basic procedures--project description, descriptive methodologies, charting techniques, information requirements, etc.); (5) personnel evaluation and in-service training (teacher performance and attitudes; evaluation of administrator training; means of assessing performance in-service training sessions). (JC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED152443

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A HANDBOOK ON EVALUATION FOR TITLE IV INDIAN EDUCATION ACT PROJECTS

READING SUPPLEMENT

RC010302

A HANDBOOK ON EVALUATION FOR TITLE IV INDIAN EDUCATION ACT PROJECTS

READING SUPPLEMENT

Needs Assessment

Financial Evaluation

The Use of Tests in Program Evaluation

Curriculum Evaluation

Personnel Evaluation and In-Service Training

NEEDS ASSESSMENT

NEEDS ASSESSMENT

The setting of goals and objectives for an educational endeavor is useful in targeting the focus of education efforts, and an analysis of educational needs can provide a means of bringing an ever sharper delineation of goals and objectives into being. Identification of educational needs, or needs assessment, is therefore usually the first step in the cyclical process of evaluation of the educational program, project, or activity, and will provide the baseline data with which other data are contrasted or compared.

This paper considers four aspects of the need assessment problem.

1. Definition and purpose of needs assessment.
2. Relation of needs to Indian Education Act.
3. Approaches to needs assessment: traditional and recent.
4. Analysis and use of information.

Definition and Purpose of Needs Assessment

Since needs evolve from the social system which identifies them and are continually being defined and redefined by the social system being served, they exist in direct relation to the perceptions of the wants and expectations of society. For example, if society

determines that all citizens must read at a functional level, then a program which produces functionally literate persons is mandated. Only when that need is fulfilled, and the system equipped so that it will continue to be fulfilled, should the need be redefined and displaced by another emerging need that then commands the attention of the program planner and the educational decision-maker.

Sources of information that contribute to the assessment of educational needs are important in planning for the actual needs assessment. Some efforts have sought to determine needs based solely on information from internal staff members from the educational system. Others have utilized literature reviews and school record files, making assumptions regarding needs of the school system extrapolated from those found in other systems. A third source of information has frequently been derived from the participation of community members themselves in establishing goals, identifying educational problems, setting educational objectives, as well as identifying educational needs (Bellott, 1971).

The hopes and aspirations of society for present and future generations are inexorably tied to the educational systems which support society. These aspirations are the basis for educational planning and find their way

into the goals of the educational system, either implicitly or explicitly (or both). Goals of educational organizations, however, are often too specific to serve as guide and direction for program development or so global as to be meaningless; at times they are found in unused archives, at other times not even in written form. Expressed or implied goals should provide direction for decision-making regarding the educational program and its attendant support systems.

If the educational goals of a society incorporate that society's hopes and aspirations, or its intentions regarding the education of its population, the educational needs will reflect society's shortcomings, failures, and oversights as well. Therefore once the goals have been expressed, needs can be identified by determining present status and focusing on the discrepancies. The assessment of educational needs may then be depicted as a three-stage process (Sarthory et al., 1974) in which GOALS minus CURRENT STATUS equals NEEDS (G - CS = N).

1. Establish GOALS,
2. Describe present status in terms of GOALS,
3. Analyze divergencies to derive needs.

Goals, needs, and program are related in that goals (where you want to go) and needs (gaps between where you are now and where you want to go) can suggest directions for program development and modification. As an example:

GOAL: Students should learn how to be good managers of time, money, and property.

Gap, discrepancy--NEED

WHERE YOU ARE NOW: Evidence that students do not understand interest rates and installment buying.

This need might suggest modifications in a variety of school programs (math, home economics, vocational agriculture, social studies) so that students <u>can</u> understand interest rates and installment buying. Or it might suggest that a new program ought to be developed to deal with consumer behavior.

Plans for the effective operation of programs directed toward particular change objectives must, to an important extent, begin with the assessment of needs. In a bureaucratic setting, however, the difficulties encountered in discerning needs are often compounded by the time framework in which needs are perceived, as well as the complexity of any given system.

Need may be said to occur when what is happening is below that which is expected. While this definition may be adequate in situations where an innovative program is being established in an area not previously addressed, or needs are of such numbers and/or magnitude that they may not all be served simultaneously, the definition is excessively narrow for other situations or for many organizations. That is, needs not only exist because of present, obvious absence of services or goods. They may also be the result of covert, latent lacks of services or goods which arose at a prior time. And they may arise as a future need for services or goods, i.e., one for which an awareness does not develop until some time after other conditions have materialized.

Participation in most federally funded programs requires, however, that an effort be made to assess the needs for which the program offers relief or improvement.

Under the requirement to conduct a "needs assessment," the applicant agency is therefore faced with determining how perceived present and future needs will be identified, measured, and given priority.

Because societal changes, changes in the aspirations of clientele, changes in laws and regulations, and a multitude of other factors influence the needs (both real and perceived) and goals of any system, the educational needs of a complex school system should be under constant review. Yet in order to maintain a sense of stability and set program direction, each system requires stated goals and plans for ensuring that these goals are pursued and implemented in predetermined priority.

Relation of Needs to Indian Education Act

Some directions that school agencies could follow in applying for funds for meeting the special educational needs of Indians are set forth in the Indian Education Act of 1972 (I.E.A.) and related documents. Some specifically mentioned are: providing for devleopment of Indian education programs in general; the provision of cultural heritage programs; provision of such supplemental academic courses as reading and math; development of teacher training programs; enhancement of Indian self-concept;

provision or expansion of counseling and guidance for Indians; vocational training of Indian students; and solution of health problems affecting education of Indians among others.

Limitations and proscriptions for school agency eligibility are included in the Act and affect the implementation of projects under the program. For example, under Part A of the I.E.A., funding is available only to local education agencies (LEAs) unless other agencies are specifically designated by the Commissioner. The Act also specifies that there must be ten Indian pupils within the LEA, or else at least 50 percent of the school population must be Indian in order to be eligible for IEA funding; another provision requires that in the planning and implementation of the program, an Indian advisory committee, at least 50 percent of whom are Indian parents, must participate.

The guidelines, limitations, and prescribed provisions of the I.E.A. are useful in program planning. They are directional; they set program parameters; and they give indications as to the kinds of activities that are looked upon with favor (or without bias). However, oftentimes these are used as the primary focus of a proposal. In such cases, guideline identification of needs to be addressed with IEA funds determine the proposed activities

rather than priorities which have been identified by the local community and agency. Less meaningful projects result because they are geared to fit the provisions of the Act rather than using those provisions to fulfill the needs of a particular LEA's Indian population. Extracting a description of needs should not be proscribed and prescribed by the I.E.A.: the most important need for a given community might not even appear among the needs identified in the language of the Act.

The Act states that the LEA must consider the inclusion of activities that build upon and support the heritage, traditions, and lifestyle of the community being served, in joint consultation with the Indian community and the parent committee. Additionally, it provides for consideration of instructional or supportive services, activities, and experiences.

Practical considerations in the development of a proposed project include the availability of funds and the degree of conformity to program goals demanded by the federal government. These considerations, however, should not impinge on the planning process; indeed, they should not even be involved in the initial planning. The first step, the assessment of needs--their identification, measurement, and priority--should not be limited by the

language of IEA. The various limitations of the Act can receive ample consideration after needs have been identified. Some provisions of the Act may actually preclude the most pressing needs from being addressed by IEA funding. But this will not prevent those needs from receiving the identification required for relief from other sources, if not from IEA.

Thus, when a school system is applying for funding, it should first ascertain the needs of the LEA related to its Indian population. It should then rank order those needs, beginning with "most acute." If the school system's project is not able to receive sufficient funding to meet the most pressing needs, the LEA would still list these needs, acknowledging that highest priorities may not be met under current guidelines. Objectives addressed to lower-priority, but lower-cost, needs would then follow. Accordingly, the needs assessment will have been carried out in a manner accurately reflecting the community under consideration. OE, legislators, etc., may thereby be informed of urgent needs not being met under existing conditions of the IEA. In this manner, the discrepancy that often occurs between needs and objectives will have been presented.

There is one point where the Act is specific to

LEAs regarding needs assessment: Indian participation at all levels of project planning and implementation is mandated. Thus, any procedures for needs assessment must be designed to include Indian participants. This means that several techniques of needs assessment that might be acceptable with other populations might be inappropriate or unsatisfactory. For instance, decisions made solely by school system administrators that did not involve community input, e.g., a survey or group meeting involving only administrators and staff, would disregard significant expression of opinion by Indian persons. An analysis of Indian students' school records and socioeconomic data would be insufficient, even though such documents provide important back-up data. In this situation, direct consultation with members of the Indian community is absent, and emphasis on other needs, not adequately reflected in the records and perhaps deemed more important, would not be presented.

Approaches to Needs Assessment

This section describes approaches to needs assessment that are well known to some school administrators and might appear to be "traditional" methods of needs assessment. They do not differ from traditional techniques, except that each is presented in the context of its use

for providing input from the Indian community. Subsequently, some of the newer approaches to needs assessment developed by the Center for the Study of Evaluation at U.C.L.A. and a study group of Phi Delta Kappa are explained. Examples of types of questionnaires for these latter approaches are also included.

Descriptions of five traditional approaches follow: (1) the survey, (2) the group meeting, (3) the Delphi Technique, (4) the use of a visiting expert or panel, and (5) the use of official records. In the presentation of each approach, attention will be given to describing its characteristics, strengths, and weaknesses. The choices regarding approaches to be employed and population sampling techniques used should be made by deciding if the cost (in resources and data quality) can be justified in terms of the alternatives and the practical constraints that exist.

Survey

Questions concerning evaluation of existing services and the need for new ones can be asked through a technique involving the social survey approach. This approach has probably been the most frequently employed technique where any attempt at systematic needs assessment has been made: administrators have sent questionnaires to or interviewed

staff members; they may have interviewed or sent questionnaries to community leaders, citizens, or parents of school children; or they have have attempted to survey all of these populations. Questionnaries may be sent to all persons in a particular group, to a sample, to leaders of particular groups, or may be distributed to as many places of communication as possible in hopes that as many persons as possible will complete and return them.

Each variation on the survey has its weaknesses and strengths which cannot be completely delineated here. Instead, the common aspects will be treated generally.

Population Sampling. The kinds of groups selected to respond to the survey will determine the kind of information derived. If only Indian members of the community were surveyed, one kind of information would be obtained. If only staff members were surveyed, another kind of information might be obtained. If only civic leaders were surveyed, still a third kind of information would be obtained. If only Indian students were surveyed, a fourth kind of information would be obtained, and so on. Ideally, the broader the base of the population sample, the more diverse the information about needs and (presumably) the more clearly will it reflect the community it represents.

The advantage of a broad base is that its multiple

vantage points may provide a more accurate picture of needs. The disadvantage of the broad base is procedural: the greater the number of viewpoints included the greater the time and money required to ascertain and analyze them. In any specific situation, a synthesis by the school system will have to be made

Kind of Questions. Two forms of questions exist for obtaining information. The closed-end question provides the largest possible number of plausible answers from which a respondent can choose. The open-ended question does not provide choices for the respondent but provides opportunity to discuss, limited only by the context of the question.

The advantage of the closed-end question is that the answers are easier to compile and analyze when the survey is completed; the number of answers in particular categories can simply be summed. Its disadvantage is that it may unjustly constrain the respondent to answer in a particular way (raising ethical issues). Moreoever, the investigator may not have provided the fullest range of possible answers. Thus, the survey may yield less information by the use of closed-end items in the questionnaire than would have been garnered from unstructured questions.

Open-ended items, on the other hand, enable exploration in a realm where the parameters are less visible, and they offer respondents more opportunity to express opinions in his/her own way. (Some respondents, in fact, will not respond to closed-ended questions because, for a variety of reasons, they consider them "insulting.")

Open-ended questions also have disadvantages, i.e., they require more time to administer than the closed-end questions; if not adequately stated, they may result in a greater range of answers than intended; answers are difficult to record in manageable form, particularly if the respondent is verbose; analysis of collected data is more difficult: categories must be generated from the data and patterns discerned from a large amount of verbiage; the process can be very time-consuming and may require complex analytical skills; and possible lack of response may result because of respondent effort involved.

Kind of Survey. As alluded to above, a survey takes the form of either interviews or questionnaires. Both methods offer disadvantages and advantages.

The questionnaire is generally cheaper than interviews, the cost of postage being less than the salary of an interviewer. Moreover, a larger number of people

can typically be surveyed in a short period of time; a large team of interviewers would be necessary to contact a large group of people in a short period of time. Also advantageous is the fact that possible bias from interviewer influence is removed in the questionnaire method.

One of the prime disadvantages of the questionnaire approach is the possibility of significant nonresponse rates occasioned by unreturned questionnaires. When response rates are low, one must then determine whether the nonrespondents are similar enough to the respondents to permit meaningful conclusions. If subgroups of the community are represented by disparate response rates (such as parents with low response rates and staff with high response rates), conclusions about needs (ascertained from both parents and staff) would be highly questionable and the value of the survey thereby reduced. Some persons, either because of the phrasing of the question or because of their own comprehension limitations, may not be able to respond in a manner satisfactory either to themselves or to the interviewer; or they may intentionally give nonsensical or sarcastic answers. The value of the survey is reduced in either way. An additional disadvantage in any type of written survey is that it assumes a level of literacy (reading, understanding, and responding in writing) which may not exist.

Some of these problems are resolved by the interview. The interviewer adds a personal dimension to the survey approach that may facilitate depth and clarity of response not possible with a questionnaire format.

A major disadvantage of the interview is that it is more expensive in terms of time and money, and "respondent bias" may occur. For example, in situations where a power differential between interviewer and the respondent exists and the questionnaire touches upon that power differential, the possibility for bias is increased. With regard to IEA programs, an Anglo interviewer asking the questions of an Indian respondent, or an Indian interviewer asking questions of an Anglo respondent, might receive an answer the respondent thinks the interviewer would like to hear rather than a candid answer. This situation can be avoided to some extent by assuring ethnic match between respondent and interviewer if questions touch upon ethnic differences. Although close monitoring and training of the interviewer can somewhat alleviate problems of personality differences between interviewer and respondent, insufficient recording of answers by the interviewer, and lack of professional integrity they cannot be totally eliminated.

The Group Meeting Approach

In addition to, or in place of the survey, some

school systems use group meetings to decide need priorities -- a method has the advantage of direct interaction in which open conflict of viewpoints can emerge and move toward resolution.

This approach also has disadvantages: viewpoints of passive or nonvocal persons are often omitted from consideration; cultural and/or other differences among those attending may result in nonparticipation by one group or domination by another group; possible vantage points are limited by those in attendance; and systematic analysis and/or derivation of needs may be lost via group process.

Frequently these group meetings are rather unstructured affairs in which participants contribute their ideas until some sort of consensus is reached. Several structured group meeting techniques, however, have been developed. The four limitations just described still hold in the use of these techniques, but they may be guarded against to some extent if leaders of the meeting are aware of them and employ positive group dynamics tactics to counteract them.

Delphi Technique

The Delphi technique (Helmer, 1967), developed by the Rand Corporation, has been useful in eliciting opinions without personal contact or face-to-face confrontation.

The process uses a series of issue questions which are repeatedly circulated and recycled with feedback to members of a panel.

Advantages of the Delphi Technique lie primarily in the lack of committee compromises which often reflect spurious influences or intimidations. Focus is on examination of an individual's position and bias for it; the entrapment of a committee member's reluctance to abandon a position or opinion he has publicly professed and then feels compelled to support is thus avoided.

Disadvantages include some of those accompanying any method of inquiry that does not elicit open-ended responses, i.e., a dependent on well-formulated questions and difficulty in administering to a larger number of participants.

The Delphi Technique is recommended as a process for bringing about a convergence of opinion, especially if a limited number of participants (a selected sample or representative group) is employed as a responding panel. The studies listed below can be useful in understanding the process and its application:

Frederick R. Cyphert and Walter L. Gant,
"The Delphi Technique: A Case Study,"
Phi Delta Kappa, LII (Jan. 1971), 272-73.

Robert C. Judd, "Delphi Method: Computerized 'Oracle' Accelerates Consensus Formation, "College and University Business, XLIX (Sept., 1970), 30-34.

Olaf Helmer, "Systematic Use of Opinions," The Rand Corp., paper presented at the AFAG Board meeting (Nov., 1967).

Panel Review

A visiting panel or expert(s) to conduct needs assessment may be used, and, in fact, has been the approach of some school systems to provide input which the school system perceives is lacking. The visitor with particular knowledge in an area may also be asked to offer a neutral viewpoint in a situation wrought with controversy, or may conduct surveys or meetings to determine needs. But the prime expertise of persons not directly involved with a particular project lies in their ability to offer advice on the basis of previous experience and special knowledge.

On the other hand, the disadvantages of experts include: limitation of perspective; lack of knowledge or rapport with the conditions of the particular community; and the risk that experts, or any outside person making recommendations, will probably not adequately represent all community views.

Records and Archival Data

Under conditions specified by the IEA, the use of school records, test scores, and socioeconomic data should not be the sole means of assessing needs because such procedures do not allow for Indian input. As sensitizing mechanisms or support data, however, they may be extremely valuable sources.

Weaknesses of these data lie primarily in their inherent incompleteness: school records may not exist regarding a specific need that appears to exist; or records may not exist that would sensitize one to a latent or heretofore unperceived need; student/family scoioeconomic history when used alone may be misleading because of failure to account for lack of program elements in assessing needs.

The strengths of this method are: it affords "objective" information about potential areas of needs; it provides information that has already been collected and is therefore inexpensive; hard data, such as achievement scores, may be easier to report, but care must be exercised in their use for needs assessment purposes.

All methods have strengths and weaknesses, and each school system should select the method that best suits its purposes, organization, population, and resources. Limitations of the time and resources often become the most important considerations in making this selection.

Recent Trends in Needs Assessment

The UCLA Center for the Study of Evaluation Symposium on Needs Assessment examined the uses, procedures, and methodology appropriate for the tasks. The essense of the symposium report (May, 1971) is the basis of the outline for needs assessment methodology provided here. According to the report, the purposes of conducting needs assessment are:

1. To determine which needs have the highest priority.
2. To justify focusing attention on some needs and not on others.
3. To provide baseline information against which to assess subsequent changes.

Further, all needs assessments should include these four basic activities.

1. Listing the full range of possible goals (or objectives) that might be involved in the needs assessment.
2. Determining the relative importance of the goals (or objectives).
3. Assessing the degree to which the important goals (or objectives) are being achieved by the program (i.e., identifying discrepancies between desired and actual performance).
4. Determining which of the discrepancies between present and desired performance are the ones most important to correct.

In attempting to derive a full range of educational goals and performance objectives and to involve the community

in the selection of these goals, the Phi Delta Kappa program (Rose, et al.) provides one such procedure:

> The [PDK] program consists of a series of strategies whereby a school district may develop community-ranked educational goals and teacher-developed objectives. Provision is made for the involvement of members of the community, the professional staff and students in:
>
> 1. Ranking of educational goals in order of their importance;
>
> 2. Assessing how well current educational programs are meeting these goals; and
>
> 3. Development of program-level performance objectives by the professional staff designed to meet the priority-ranked goals.

This program establishes a game situation in which either community representatives or the entire community takes part. Large numbers of participants do not place intolerable burdens on the study staff and do improve the quality (acceptance) of the data. While the entire PDK program begins with ranking of goal priorities and carries through the design of instructional programs utilizing performance objectives, as well as providing a resource allocation procedure needed to meet these objectives, the control of the program and the choices regarding how extensively the district will participate always remain the option of the school district decision-maker. The information derived from goal-setting using

this procedure reveals how community members perceive the importance of each of the goals and where they feel the efforts of the district should be placed. Needs are thus derived from goal setting, and priorities are decided through community participation.

Three factors are revealed regarding each goal: (1) the perceived current level of performance, (2) the desired level of performance, and (3) the emphasis given by the school. The perceived needs of the school district can then be quantifiably identified by the scores attributed to each goal.

Analysis and Use

For any of the procedures described above, the scope of data analysis is determined by the complexity of the survey instrument, the size of the sample, and the complexity of needs assessment deemed necessary. In many cases a simple frequency count for each of the questions is sufficient. In others, some cross-tabulations are necessary to ascertain the perceptions of different sectors of the sample and what they believe to be the most important needs. Cross-tabulations may yield information about differences of opinion, if any, between groups. Every effort should be made to take into account the various opinions,

with special attention to needs and priorities identified by Indian students, parents, and community.

All of these factors become a part of the analytic design of the needs assessment study. Although such a study should yield information consistently obtained from specified groups of people, the process should be viewed as more than a "one-time-only" attempt to obtain information. The character of any particular needs assessment study will differ according to the design established by the investigator, and the data analysis will certainly reflect the design. However, the following analytic techniques have been found useful in dealing with multiple input sources:

1. Grouping of data by variables which are descriptive of population characteristics (e.g., ethnic group, age, sex, educational level, income level).

2. Discrepancy analysis to determine the gap between the current status of educational programming and that which the population says is expected or desired.

3. Weighting of data in such manner that each data source (or separate population) can be treated as equal to any other data source, regardless of size of sample or size of populations e.g., use of Garrett' order-of-merit conversion.

4. Use of statistical procedures based on population characteristics.

Assessment of educational needs is not only the beginning of the evaluation process, it is also the beginning of the fulfillment process--assuming that one must identify a need before one ministers to the need. Unless assessment reveals a deficit or shortcoming, no motivation exists to take action to remedy it.

The very process of needs assessment can enhance the educational system if it elicits participation by and from its community. This may well be fulfillment of another set of crucial needs arising as a spin-off from the primary reasons for participation.

REFERENCES

Bellott, F.K. Design for Tennesseee assessment and evaluation of Title III, E.S.E.A. Memphis, Tn.: Memphis State University, 1969.

Bellott, F.K. Needs Assessment: A review of local practices. Memphis, Tn.: Memphis State University, B.E.R.S., 1971.

Carter, A., Ray, L., and Bellott, F. Educational expectations survey form "A". In Longitudinal study of Memphis city schools. Report in preparation, 1975.

Carter, A., Ray, J., and Bellott, F. Educational goals form "B". In Longitudinal study of Memphis city schools. Report in preparation, 1975.

Coffing, R. T., Hodson, W. A., and Hutchinson, T. E. A needs analysis methodology for education of the handicapped--version 1 (Connecticut Project No. BB-404). North Haven, Conn.: Area Cooperative Educational Services, July 1973.

Coffing, R. T. and Hutchinson, T.E. Needs analysis methodology: A prescriptive set of rules and procedures for identifying, defining, and measuring needs. In Methodologies under development. Symposium presented at the meeting of the American Educational Research Association, Chicago, 1974.

Helmer, Olaf. The use of the Delphi technique in problems of educational innovation. Santa Monica, California: Rand Corporation, December, 1966.

Lipman, J. M., Gregg. T.T., McIssac, D. N., and Morrow, R.G. Wisconsin educational needs assessment study. Madison, Wis.: Department of Public Instruction, June 1969.

National Advisory Council on Indian Education. The Indian Education Act of 1972. Washington, D.C.

Rose, B.K., Balasek, Jr., Kelleher, J., and Nelken, I. A programmed course for the writing of performance objectives. Chico, Calif.: Northern California Program Development Center, 1971.

Rose, B.K., Balasek, J., Kelleher, J., and Nelken, I. Educational goals and objectives: A model program for community and professional involvement. Chico, Calif.,: Northern California Program Development Center. N.D.

Sarthory, J. A., Douglas, J. R., Holloway, W., and Van Meter, E. J. Project Kansas 76: Preparation of educational personnel for new leadership roles in the public schools. Topeka, Kan.: Kansas State Department of Education, June 1974.

Slotnick, H. B. National assessment of educational progress, report 8, writing: National results -- writing mechanics. Washington, D.C.: U. S. Government Printing Office, 1972.

UCLA Center for Study of Evaluation, Report. Los Angeles: University of California at Los Angeles, 1971.

Van Dalen, D.B. Understanding educational research. New York: McGraw-Hill, 1966.

Walker, H. and Lev, J. Statistical inference. New York: Holt, Rinehart, and Winston, 1953.

FINANCIAL EVALUATION

FINANCIAL INFORMATION

Financial information for the assessment of service programs and projects is derived by collecting and processing data pertaining to the expenditure and utilization of financial resources--money. The process by which this data is compiled, translated into required or useful form, and related to the goals and objectives of human services programs in general and Indian Education Act projects in particular will be presented in this paper by identifying and describing:

(1) Financial information requirements for programs and projects;

(2) Procedures for selecting the type of data to be collected to meet these informational requirements;

(3) Procedures for collecting and processing data; and,

(4) Recommendations and cautions regarding the use of financial information in program decision-making.

This presentation is intended to be general in nature and the specific procedures described are to be applied or discarded in the light of the unique requirements, capabilities, and resources of the various IEA projects.

Financial Information and the IEA

Before describing any procedures, it may be well to review briefly the requirements for financial information at various levels of IEA program management and the ways in which this information may be used.

Generally financial information derived from human services programs is used for budget control, for resource allocation or reallocation, and for informing various publics and governmental agencies. In regard to IEA programs, the Office of Indian Education (OIE) uses such information to ensure that funds are employed as intended by the authorizing legislation and to evaluate the effectiveness with which program funds are used.

All IEA projects are required to provide a common set of financial data to OIE which will explicitly state nature and purpose, proposed use, or the rationale for request. For example, if OIE receives requests for financial information from Congress and the Office of Management and Budget, these requests must be honored for administrative and political reasons. Usually this information pertains to the number of people directly and indirectly served by a project and the total project budget, and it is used for either justifying or criticizing the programs sponsored by OIE.

Cost data, which all projects are required to submit to OIE, should, at the same time, be useful to local project management. In this way, local project management is given

the opportunity to review and revise lists of data itself, as well as the procedures for data recording, prior to the time of official OIE publication. The final format, as well as procedures for collecting financial data, will take into consideration the needs of both the local projects and the federal government.

Projects are encouraged to collect, process, and use information which may not be requested by OIE but which may prove useful in the planning, development, operation, and enhancement of specific projects and activities. Such additional uses will be so designated and discussed here.

Identifying Financial Information Needs

Financial information, by definition, includes "dollar" data and other facts required in analyzing and making this data relevant, of which the cost-per-person-served by the program is a good example. To compute cost-per-person-served, two further types of data are required: (1) the cost of the program, and (2) the number of persons served by the program. Financial information requirements may be identified by using one of several procedures.

Review of Project Objectives

Each project has a set of objectives to accomplish, and an objective specifies, for example, that "each student entitled to project services will be examined for the purpose of identifying hearing disabilities." To measure the degree to which this objective is accomplished, two categories of information

are needed: (1) the number of students entitled to program services, and (2) the number of students examined for hearing disabilities.

Some objectives, however, do not state such obvious informational factors. For example, in order to measure the stated objective that "each native American student will have the optional opportunity for counseling services provided by a certified Native American school counselor; the counselor-to-student ration will not exceed 1:200," several obvious pieces of data are required. How many native American students are in the school? How many certified American counselors have been employed? However, the objective as stated also uses the terms "optional opportunity" and "counseling services" without defining explicitly what is meant. Such specific information requirements are difficult to identify except at a very gross level, e.g., what types of counseling services are being rendered and how are services optionally provided? To identify more explicitly this needed information, still more information is needed before the relation between cost of program, implementation of program, and evaluation of program can be established.

Review of Project Plans

Project proposals submitted to OIE are described in narrative form, and in some cases include an explicit listing

of activities, along with a PERT chart denoting when and in what sequence the project activities will be carried out. This narrative and listing of activities should be reviewed in identifying informational needs. In the example referred to above concerning "optional opportunity for counseling services," a review of the narrative should indicate the meaning of the term "optional opportunity." If, for example, the narrative indicated that students would not be required to rap with the counselor, but that each student could if he or she so opted, then additional information requirements became more obvious. How many students opted to rap with the counselor? How often? For how long? Were students turned away and not served by the counselor? For what reasons? This type of data then provides information on the extent of the counseling service.

Similarly, the term "counseling service" may be defined in the narrative with a listing of information requirements pertaining to the quality of service extended. For example, if "counseling service" includes assistance in identifying and obtaining a tutor for a student who is experiencing academic difficulties, then the number of cases where tutors were obtained and, more importantly, the number of cases where the service tended to solve or alleviate the problem, are both examples of information required in evaluating the project.

Project activities specified in the proposal should be reviewed to determine the types of information to be collected and processed if the project is to be effectively evaluated in terms of extent and quality of program implementation.

A review of objectives and narratives, however, does not necessarily occur only once. In the time from proposal writing to the actual start of a project, major changes can occur in objectives and/or strategies for achieving objectives, and, as the project develops, there will be operational changes. Therefore, any change in objectives and/or activities should automatically trigger a review of information for financial data requirements.

Review of Selected Cost Accounting/Evaluation Procedures

Most of the needed information identified in reviewing project objectives and plans tends to pertain to other than dollar data, such as: the number served, the extent of services, and the quality of services. A review of some selected cost accounting and cost evaluation procedures will assist in identifying most dollar information requirements.

A simple accounting system utilizing the budget guidelines and some or all of these categories, may be employed to meet these government requirements. The federal government requires periodic reports on the expenditure of project funds and a final fund expenditure report prepared at the

end of the project. Projects prepare proposed budgets according to guidelines and, generally, the same budget categories used in preparing proposals are required in submitting expenditure information to the federal government as a requisite for receiving funds. (Budget categories usually include: Personnel, Fringe Benefits, Travel, Equipment, Supplies, Contract Services, Construction, Indirect Charges, and "Other.")

Another system uses budget categories and identifies the amount of money spent within the various project activities. This system has an advantage over the simpler system just noted in that it takes into account the limited resources of projects and, consequently, the care that must be taken in allocating resources to the various operations of a project. Referring again to the examples used earlier, if a project had the two objectives of providing hearing tests and counseling, the project staff and advisory committee probably would be responsible for allocating resources to those activities. Thus, the number of students to be tested for hearing disabilities is partly determined by the cost of testing a child. Similarly, because a counselor without counseling materials probably will not be very effective, the amount of money allocated for purchase of these materials also becomes a part of the cost of the service.

This process of particularlizing and limiting cost to

the various objectives or project functions can be extremely helpful in the allocation and control of funds, and the matching of cost-budgeted items (or expenditures) with specific project functions is essential if the project is interested in answers to questions such as, "What was the cost per student served?" If all personnel costs, for example, for this project were lumped together, then the cost of personnel for students served by the counselor would have to be added to the cost of personnel for the number of students tested for hearing disabilities. The resulting information about cost-per-student served is thus less usable as data in making decisions about future resource allocations.

Therefore, project questions concerning cost related to each objective or function can be answered if the accounting system provides for identifying cost within each project component. Therefore, with regard to counseling and hearing test functions, personnel cost will be recorded as: (1) personnel cost for a certified native American counselor; (2) total personnel cost for the counseling function; (3) personnel cost for the hearing-testing component; (4) part-time clerk responsible for filling out information forms for students tested, etc.

The selection of a cost-accounting procedure which provides more data than explicitly required by federal guidelines

should be based on project staff and advisory committee desire for additional information which will contribute to better allocation and spending of resources, and the identification of this needed information comes from the review of the potential accounting system rather than from review of a system already selected.

A third accounting system that could be selected by a project is an extension of the second. For example, if a project is using two different approaches to achieve the same objective, then the accounting system should be designed to identify money spent for each approach, e.g., Indian language taught to one group of students by a certified teacher and to another group by a member of the community. Because the personnel cost would probably be different for each approach, the accounting category, personnel, would be divided by project function. These objectives or functions (hearing test, counseling service, and language instruction) can be further divided by the two approaches, teacher instructed and community instructed.

By collecting cost data on the two different approaches and by collecting some common type of performance scores for each approach, project personnel (staff and advisory committee members) will be able to review the cost-effectiveness of the two approaches. For example, if the teacher-instructed

group scores only slightly higher on the performance test than the group taught by the community member, and if the teacher cost was 25 percent more than the community-member cost, then the project personnel may decide that the community-member-instructed approach is more "cost effective" and for the next year only that approach should be used.

This cost-effective approach to accounting can be expanded to include data for categories other than cost and performance. In the case of using two different approaches for achieving the same objective by the use of two different types of teacher, if the teacher responsible for teaching the language was not a member of the Native American community familiar with that particular language, and if the teacher had not been sufficiently trained in the language, then the teacher-instructed approach may not have been implemented well, and this would have affected the performance of the students. Hence, the "test" for deciding between the two approaches may have been loaded in favor of the community-member-instructed approach, and any decision to continue one or both of the approaches would have to take this into consideration. Therefore, information pertaining to the <u>implementation</u> of a project activity may be recorded and used in making judgments on cost-effectiveness.

Also useful in reaching cost-effective decisions are data on the characteristics of people being served by the

project. Using the example of language instruction, students who have a high familiarity with the language prior to instruction may perform better when taught by a community member than by a certified teacher, while students with a low familiarity with the language prior to instruction may perform better when taught by a certified teacher. Cost-effective data then may suggest that both approaches be continued and assignment of students be based on familiarity with the language. Should a project decide to utilize such cost-effective data, information to be collected and processed would then include:

(1) The cost of approaches

(2) Effectiveness of the approaches

(3) Implementation of the approaches

(4) Student characteristics data.

Although data on categories (3) and (4) are not essential to a cost-effective analysis, and such an analysis may be performed if only data on categories (1) and (2) are available, data on implementation of approaches and student characteristics do provide additional information which is very useful in making operational decisions.

In summary, this cost-effective approach, comprising several levels of complexity, provides a third level of cost accounting that may be considered by project personnel in

selecting a financial accounting procedure, requiring the collection of performance or achievement data, and may also require the collection of implementation and/or student-characteristics data if the project opts for the more complex analysis of project effectiveness in addition to cost effectiveness.

The following table summarizes the types of information that would be required to implement the three accounting procedures reviewed.

	Cost by budget categories	Cost by budget categories subdivided for objectives and/or components	Cost by budget categories subdivided by approaches
Minimal cost accounting procedure	X		
Cost accounting procedure which relates cost to objectives and/or project components	X	X	
Cost-effective procedures	X	X	X

Review of Information Requirements with OIE Personnel

The final procedure suggested for identifying needed financial information is to solicit suggestions from OIE personnel. Their experience in responding to questions raised by Administration officials and Congress should be tapped while trying to identify information which projects should have available for both local and national audiences. Three categories of information which may be suggested are:

(1) Titles, numbers and percentage of Native American staff personnel and amount and percentage of money paid to these staff members.

(2) Percentage of total budget expended for project administration.

(3) Types and amounts of in-kind resources available to the project and sources of these funds.

Selecting Data to be Collected

Selecting data to be collected and processed follows from the formulation of information needs. Several factors must be considered in selecting the data:

(1) Level of information requirements decided by the project.

(2) Availability of data.

(3) Availability of staff and resources to collect and process the data.

(4) Potential for collecting accurate data.

(5) The multi-use or limited use of data.

Level of Information Requirements

The minimal level of information required and the data collected to fulfill this requirement will be established by OIE, following meetings with project directors. Data to be collected in addition to this minimal requirement will be determined by each individual project as part of the project evaluation. (Information requirements identified in conjunction with OIE personnel should also be reviewed.)

If the project is interested in cost-effective analysis, data on the various levels discussed in the description of this procedure will have to be collected. Prior to this, such questions as the level of project implementation desired must be answered to determine the level of informatio on which the project will base decisions. The primary rule in determining the level of information to be collected and processed by a project is to plan to collect and process only data which will be used in making decisions about project [illegible] and operations.

[illegible]he minimal level, which concerns OIE required data, is one option. If that option is exercised, then the following factors need not be considered in selecting data to be collected, but these factors may prove helpful in developing data collection and processing procedures.

Availability of Data

After determining the level of information required by the project, the availability of data to meet the information

need should be investigated; it cannot be taken for granted. For example, if a project should opt to implement a cost-effective accounting procedure, will the performance measures be available to project staff? In the case where a tutorial service is provided by a community group to students, would the test data from this service be available to project personnel, or would the project have to perform additional testing?

Another question concerns timing even if data are available. The school system may agree to provide the data when they are available, but they may not be available until after the need for the data has passed. For example, if plans were made to use standardized achievement scores in evaluating the effectiveness of the counseling component of a project, the results must be available at the time of organizational personnel, and budget planning. Decisions regarding continuation of program components are also heavily dependent on a time factor in the matter of adequate modification of personnel.

Availability of Staff and/or Resources

The amount of time and therefore staff required to collect and process data is another factor to be considered. Some data may be available in a form which would require no additional processing, or take only a telephone call to acquire; staff time to collect this data would thus be minimal. Other data collection may be more time-consuming. For

example, how would data be collected to determine how well the native-language-instruction component of a project was being implemented? The use of trained classroom observers to collect the data could be very time-consuming, and time can be translated to cost; collection of data would thus be a major cost item. Further, is one observation in a classroom sufficient, or does the observer need to visit the same classroom at least once a month during the school year? Again, time is translated into the expenditure of resources, and the question of the need for collecting these data must be raised. In a situation where language instruction is not being implemented and operated as planned, then all the resources expended for this component may be lost. The expenditure of funds to collect data to aid in implementing the component may therefore be an example of resources well spent.

Potential for Collecting Accurate Data

Some individuals respond to interviewers with answers he/she believes the interviewer would like to hear, rather than that with candid answers. Some students may not take a test seriously if they do not consider it relevant. A clerk may record that all children have had their hearing tested rather than checking all records to ensure that no one has been missed. A school counselor may spend more time performing administrative jobs than providing counseling

services. Because these instances of possible ways in which inaccurate data can be recorded represent only samples from an endless list, the probability of collecting accurate information must be taken into consideration in determining which data will be selected for collection.

Project budget and the record of expenditures can provide sources for cost data which will meet the dollar information requirements of project evaluation. The accuracy of these data will, however, depend on the monetary control exercised by project staff; loose control may result in questionable cost data.

Data collected by counting things (student tests for hearing disabilities, students serviced by a counselor, students taught a native language) may be controlled for accuracy by performing quick audits of the data: having a different staff member check school records for the number of students eligible for a service; recounting the project's records; checking with a few students to determine if they have received services as recorded.

The accuracy of achievement and performance data is, in turn, also dependent on a variety of factors, such as

how well the test, or instrument, measures what is being taught and how well the instrument is administered to students.

Data on student characteristics may be either easy to check for accuracy (counting the number of students) or very difficult to check (specifying attitudes of individual students). Some student characteristics data may be a personal nature, e.g., the range of family income, and these are collected by checking the number of students participating in a free or reduced-price lunch program. However, any attempt to audit this information would be an incursion on personal rights.

Project implementation data is normally collected through either interviews or observation, and accuracy depends largely on the training and skills of the interviewer and/or observer, as well as on how effectively the interviews and observations are planned. A check for accuracy may be performed by two or more interviewers and observers working with the individuals being served by the same project components. Students need not be interviewed twice, but two people interviewing students from the same class should be collecting comparable data.

A complete audit of all data collected would be time-consuming and costly, and is not required. Projects should assume the accuracy of most of their data and only audit where the greatest chance of "bad" data may occur.

Limited Use or Multi-Use of Data

In performing a cost-effective analysis of, say, two instructional approaches, implementation data may be easy to collect for one approach and very difficult, costly, or, in some cases, unfeasible to collect for the second approach. This would then limit the use of the data collected on the first approach because comparable data necessary for cost-effective analysis would not be available. The limited use of the data, then, should make its collection lower in priority, or inadvisable.

Some data may have multiple uses. The number of persons served may be used in (1) reporting simply the number served, (2) computing and reporting the cost per person served, and (3) performing a cost-effective analysis on two service approaches. A high priority would therefore be given to the collection of this type of data.

All of these factors must be considered in determining which data will be collected. Most projects will not have the resources to collect all the desired data, and priorities will have to be set. Decisions as to what data to collect and not to collect will be governed by two prime factors: the data requirements of OIE and the resources available.

Collecting and Processing Data

Essentially, four different methods of collecting data are in use:

(1) Using a questionnaire completed by the responder.

(2) Interviewing staff and individuals served by the project and recording the information in some standard format.

(3) Observing project activities and recording the information in some standard format.

(4) Recording project activities on video and/or audiotape.

In considering the advantages and disadvantages of these four techniques, no simple formula for selecting the best procedure presents itself. Each must be weighed in terms of attaining a workable balance of goals and resources for each program.

Once the data have been collected, some degree of processing is required before they are usable for decision-making, and this processing may range from simply totalling numbers to complete statistical analysis.

The analysis of most financial information is not complex. It normally consists of totalling sums and constructing matrices which describe the relationship between dollar data and related information. A minimal cost accounting

system does not require the amount of nondollar data which the cost-effective model requires. Processing of data for the minimal cost accounting system consists mostly of the recording of cost data in the appropriate budget categories and comparing budget with actual expenditures.

The processing of data required by the cost accounting system, which relates costs to project objectives and/or components, is very similar to the processing required by the minimal cost accounting system. The complexity is determined by the number of budget subcategories.

The process of collecting financial data becomes complex if a project selects an accounting system which provides for cost-effective analysis because processing then involves an increased number of budget subcategories. Again, summing of financial data is required and this data must be related to achievement, implementation, and information concerning student characteristics. In analyzing this data, cost per approach must be calculated and students therefore divided into two or more groups, depending upon student characteristics (e.g., by sex and family income, yielding four groups: girls/high income, girls/low income, boys/high income, and boys/low income). These groups, in turn, must be divided into students who were served according to a well-implemented plan and students who were served where the approach was not

implemented according to plan. Each of these groups then has a cost effective ratio the per-student cost of the approach divided by the average effectiveness measure for students grouped. Comparison of these groupings may then be used to compare the relative worth of the approaches.

Evaluation has its basis in information, and information is a powerful ally in accomplishing goals. Information pertaining to dollars is particularly powerful, and evaluation can provide that information. This presentation has attempted to outline material to assist projects in the development of a plan which will give them that informational power.

But because information is only as good as the data upon which it is based, care must always be taken to ensure that the data collected and processed is accurate, timely, complete, and relevant. Evaluation is a means for improving the responsiveness of projects to the individuals they are intended to serve; it is not an end for the project or program. It should always be used as an informational tool for decision-making, but should not be the determiner of what should occur.

THE USE OF TESTS IN PROGRAM EVALUATION

THE USE OF TESTS IN PROGRAM EVALUATION

Testing is a crucial part of the measurement aspect of evaluation and has often been considered synonomous with evaluation. Broader educational practices, however, now place it in a larger perspective in which the issue of what test should be used to measure the worth of a program is one of several critical aspects of decision-making. Testing, nevertheless, remains an integral part of most evaluation schemes and thus deserves renewed attention.

Although in this presentation, testing refers primarily to standardized testing and is designed to serve as an aid to Title IV personnel in the selection, application, and interpretation of standardized tests, some limitations and cautions, as well as the useful and enhancing aspects of standardized testing, are discussed. Because achievement testing is more widely used than other forms of standardized testing (such as aptitude and intelligence or personality and attitude) in Title IV programs, when practical suggestions are made, they will focus here.

Tests as evaluative tools, may be categorized as to what they measure, the data they provide, and how they are constructed to facilitate the results they attempt to provide. Within these categories, tests may be classified in at least three different ways. They may be grouped

according to their content: achievement, aptitude and intelligence. Or they may be classified according to the kind of data they provide and time at which the data are collected, i.e., immediate feedback, or formative data about an ongoing program or project, usually focused on individual students or groups and used for curriculum planning and revision. A third important classification of tests -- norm-referenced and criterion-referenced -- are structured on differing bases for rating performance by comparison.

This discussion of testing will focus on some of the questions most frequently asked by those considering the use of tests for evaluation. It is hoped that this discussion will provide the reader with suggestions about how to proceed, discussion of common testing concerns, or references for further investigation. Questions as to "right test," or "best test" are superseded by questions perhaps best described in terms of a wide-ranging concern for specificity and appropriateness:

-- What information is sought from the testing?

-- Under what conditions can this be obtained?

-- What are the characteristics of the test that fulfill these conditions?

The questions to be discussed are:

I. What are the criteria for test selection for Title IV projects?

II. How can knowledge of the specialized vocabulary of standardized tests function in maximizing their usefulness as evaluative tools?

III. What are the differences between norm-referenced tests and criterion-referenced tests.

IV. What are some limitations of standardized tests?

V. What are the major steps involved in using tests for evaluation of educational achievement?

I. What are the criteria for test selection?

In preparing their CSE Elementary School Test Evaluation, the Center for the Study of Evaluation in the Graduate School of Education at the University of California at Los Angeles used several criteria which should prove helpful for persons evaluating tests for use in Title IV projects:

- Validity (Does the test measure what it is intended to measure?)
- Appropriateness to examinees.
- Administrative usability.
- Technical excellence.

Lake, Miles, and Earle, in Measuring Human Behavior, list five criteria by which tests may be judged:

- Must give same results when used by different people.
- Must not change properties over time.
- Must be related to an important aspect of the world.
- Must be relevant to actual decisions which must be made.
- User must have confidence in results obtained.

The following additional references may be helpful for those desiring to know more about the qualities of adequate tests:

Building a School Testing Program, 1969 Impression, R. H. Bauerfeind. Houghton Mifflin, Boston, 1969.

Standardized Tests in Education, W. A. Mehrens, and I. J. Lehmann. Holt, Rinehart & Winston, New York, 1969.

Testing (Guidance Monograph Series, III) S. C. Stone and B. Shertzer (Eds.). Houghton Mifflin, Boston, 1968.

A number of factors occur in the test selection process:

- The purpose for testing (diagnosis, placement, or assessment of progress) must be ascertained;
- The content area, or learning objectives, or domain to be tested must be clear;
- Awareness of various tests available in the particular areas of focus;
- Narrowing the field and then reading or obtaining reviews of thes tests;
- Review of the test in its entirety to see that it seems to do what it claims to do.

To help in this process, discussion with others who have been in similar situations (other Title IV project personnel, for example) or a tests and measurement expert may prove useful. The first few steps of the process should probably be carried out by a local project person. Unless

there is knowledge of a clear, comprehensive, and specific set of objectives, an outsider, even if she/he is an expert, will only be guessing at the program's intents. The search itself can be aided with the use of sources and references some of which are presented at the end of this paper, grouped according to:

A. Books containing inventories and reviews of tests.

B. Books describing the general field of standardized testing.

C. Major test publishers.

II. How can knowledge of the specialized vocabulary of standardized tests function in maximizing their usefulness as evaluation tools?

Validity: How well does the test measure what it intends to do?

- The content validity of a test is a reflection of how closely the test contents match up with the actual subject area being measured. For instance, if a history course focused on personalities in history and a final test covered dates in history, that test would have poor content validity in this instance.

- Face validity refers to how "good" or useful the test appears to the user. This appraisal will affect the confidence she/he places in it, and thus its value as an aid to decision-making.

- Related to face validity is decision-maker validity, which refers to the confidence and subsequent use a person gives to the results of a test, as well as to whether the test helps the decision-maker make better choices than would have occurred by chance.

- Construct validity asks if the test correlates with the psychological characteristics theoretically linked to the construct being measured. Although construct validity is generally not of concern to program evaluators, it might be useful if testing is investigating

psychological or attitudinal factors which cannot be tied directly to observables.

- Predictive and concurrent validity both refer to how well the test correlates with other measures. In predictive validity studies, one has to wait for some future event or events to occur. When checking concurrent validity, an immediate comparison can be made.

Reliability: This term refers to the consistency of the test. Most often an indication of reliability is given as a correlation coefficient of which there are three main types.

- Coefficient of stability tells how well the test measures over time. Does it consistently measure what it is designed to, testing after testing?
- Coefficient of internal stability. How well do the various parts of the test correlate with each other?
- Coefficient of equivalence. Many tests have different forms; when the results of these different forms are compared, the coefficient of equivalence is obtained.

Norm: A national or local standard group to which the test population is compared that gives a test score its meaning. Most standardized tests are norm referenced. The norms to which standardized tests are referred are often criticized as "culture-bound" white, middle-class norms. When any test user suspects that his or her test takers differ from the norm, then the results should be interpreted with great caution.

Gain Scores: These reflect the impact of a program, or simply students' maturation between one testing situation and another. Caution is advised in placing a great deal of emphasis on these, especially in certain cases: a phenomenon known as the "regression effect" may occur which tends to give unusally high gain scores, especially to persons scoring in the lower extreme.

Mastery Scores: This is a term, generally used with criterion-referenced tests, which refers to a cutoff score. If a person scores above this mark, one assumes proficiency; a score below this indicates inadequate performance.

Raw Scores and Converted Scores: Bearing in mind that all test scores are indirect measurements, there are various levels at which these scores will be presented. Raw scores tell the exact number of "correct" responses. Converted scores, on the other hand, cover a wider range; they are actually more accurate than exact scores because they take into account the various errors of measurement present in all testing situations.

Some common converted scores are: grade equivalents, age equivalents, T-scores, stanines, and quotients. While converted scores, which cover a range, are more accurate (or less likely to be misinterpreted) than, say, percentiles, reporting raw scores facilitates individualizing the tests. This might be done, for instance, by eliminating responses to items which are inappropriate to a particular situation, e.g. subsection on map skills if map skills are not taught.

Statistical Concepts: Mean, Median, Mode, and Standard Deviation

Mean: the arithmetic average used to describe the average score on a test, the sum of all scores is divided by the number of separate scores (N).

Mode: the most commonly occurring score; a test may have a mode of 98, but a mean of 50 if, say, 98 was the only score which occurred more than once.

Median: the middle score, halfway between the highest and lowest scores.

Standard Deviation: a representation of the range or spread of scores. Although it is not crucial to know how it is derived, it can be a most valuable statistic for test interpretation. For instance, two fourth-grade classes could have the same mean grade equivalency, say 4.3, but vary greatly if the standard deviations were different. If class A had a standard deviation (S.D.) of .2 and this score was both added and subtracted from the mean, this would signify that roughly two-thirds of all students in the class scored between 4.1 and 4.5 On the other hand, if class B had a standard deviation of 2.0, then this class would be much more widely spread out, with a range of 2.3 to 6.3 (indicating that this second class would probably require more individualized instruction.)

III. What are the differences between norm-referenced tests and criterion-referenced tests?

Both types of tests are designed to measure performance through some kind of comparison. In criterion-referenced testing, the scores are compared with specified, established standards, like ability to add two digit numbers so that each score acquires meaning as it relates to the pre-established criterion or cut-off point. In norm-referenced testing the individual's scores are compared against each other, and each score acquires meaning only as it relates to other scores.

This main difference reflects the way test items are constructed. Because purpose of norm-referenced tests is to spread people out in order to classify, place, or select, many "distractors" -- items close to the correct answer -- are used. In criterion-referenced tests, however, the most important attribute of the item is that it accurately reflects the objective. For instance, a driver's test can be considered a criterion-referenced test: it does not matter whether a certain percentage of people pass or fail, but it does matter that the skills tested are those needed for safe driving.

The current trend is toward criterion-referenced tests, and Title IV programs might be well advised to follow this trend. However, in most cases this will call for the local preparation of instruments. Many of the areas of concern of Title IV have not yet had standardized criterion-referenced tests prepared for them. To prepare local tests, item banks

are available containing questions tied to learning objectives, mostly in the cognitive domain. To do a good job of preparing appropriate criterion-referenced tests, however, expert consultants should be employed.

IV. What are some limitations of standardized tests?

Standardized tests serve many valuable purposes. They are effective for grouping and judging overall differences of programs, and they are convenient. However, precautions should be taken when employing standardized tests in an evaluation scheme. The limitations of this use relate primarily to four aspects of the educational testing process:

A. Instructional programs

B. Students (testees)

C. The test itself

D. Interpretation and utilization.

These should only be viewed as general in nature and not a comprehensive listing of all potential weaknesses.

A. Instructional Programs

There are two prime potential weaknesses of standardized tests as they relate to educational programs. In the first case, teachers often "teaching for the test" -- drilling students on those skills which are covered by standardized tests, with little regard for the appropriateness of these skills for the individuals, or for accomplishing the overall objectives of the educational program. Another, almost

opposite occurrence is found when standardized tests fail to cover the aspects of the program which were emphasized. Using inappropriate tests in this instance could easily lead to incorrect decisions about the program's merit. For example, if a language arts program stressed creative expression, but the test measured grammatical skills, the program has a high probability of failing by normative standards.

B. Students (Testees)

The general limitation of standardized tests in regard to testees is their basic assumption that learning is steady or regular throughout the nine months of the school year. Rather, learning seems to occur in cycles, as does maturation, and students may not always show a gradual, steady progression in the acquisition of skills that achievement tests assume.

The same variables that affect students in everyday situations also affect them when they take tests; they may even be exaggerated under the stress of test conditions. Physical disorders like poor eyesight or hearing can affect performance, and moods or depressions can similarly influence test results. Further, some students are "test anxious" and score poorly when under pressure. Administrators and teachers should freely share information about these limitations with parents and students when appropriate.

C. The Test Itself

1. Limitations and criticisms of intelligence tests are especially prevalent. Because intelligence is not a singular entity, but a complex attribute with many interrelated psychological and social aspects, the attempt to measure this complexity by means of a single test therefore calls into question the nature of the test itself. The continued authority accorded intelligence scores, when these have been shown to be poor indicators of career success, further complicates assessment of the value, assumptions, and expectations of testing in this area.

2. Few standardized tests measure the "higher" cognitive abilities well, like synthesis or evaluation. Likewise, few instruments are available outside the cognitive domain, such as tests of psychomotor skills or affective qualities, that are solidly grounded in theory and have been field-tested.

3. Norming groups differ between standardized tests, and thus a person taking different tests may score much differently.

4. Standardized tests may reinforce sexist or racist views through either the language or assumption of the test itself or during interpretation and utilization of the test.

D. Interpretation and Utilization

People often fail to realize that a score (like a percentile rank) is only an approximation of what the "true"

score is. Small differences between scores should not be used as the basis for decision-making. In fact, standardized tests alone should not form the basis for decision-making; standardized test results should be combined with those of other measurement strategies. Parents often tend to give too much weight to the results of standardized tests. This can be prevented by having teachers or counselors explain some of the weaknesses of standardized tests at the same time they provide the scores (as school officials are now required to do, by law).

Rhode Island uses special interpretation workshops in which specialists are called in to teachers' meetings to explain how the test results can be used (and misused). Title IV projects might benefit from a similar strategy.

V. What are the major steps involved in using tests for evaluation?

There is no single, accepted set of rules for the use of tests in evaluation. The following nine steps may or may not occur in all evaluations, and they may not occur in the same order as they are listed below. Further, they relate to the use of standardized instruments which may be different than locally prepared measures perhaps in more frequent use in Title IV evaluations. With these limitations, nine general steps are listed and described

1 -- Know the purpose for testing.

2 -- Be aware of the limitations of instruments.

3 -- Describe the important characteristics of the group to be tested.

4 -- Understand the educational objectives of the program or project.

5 -- Estimate the type and amount of resources available for testing.

6 -- Search for or develop the tests appropriate to observe what you want to see.

7 -- Select and personally review the test.

8 -- Administer the instrument (s).

9 -- Interpret results, make decisons.

1 -- Know the purpose for testing.

Within the general framework of evaluation the three major reasons for testing are:

. To select, place or classify;

. To diagnose;

. To assess progress or growth of individuals and groups.

Different persons in the same project will have different pursposes for testing. Teachers, or staff members may want to use standardized reading tests for placement, while administrators may be more concerned with obtaining a general indication of group achievement. The type of test used will thus differ according to the purpose of testing.

Some tests can serve multiple purposes. For example, one form of a math achievement test may be given in the Fall for placement and some diagnosis; that same test might be used again in the Spring for assessment of progress. (Here, however, instructional precautions must be taken to avoid "teaching for the test.")

2 -- Be aware of the limitations of standardized instruments.

The major limitations to standardized testing that have been discussed have received much publicity recently. One should bear in mind, however, that many of these weaknesses refer to the use of the results of the tests, not necessarily to the tests themselves. The test user should be aware of these possible failings not only to aid in the selection of approriate tests, but also to be able to defend the use of tests to the public. If the administrator is aware of potential faults of tests, she/he can more easily gain community support for them. The following charges against reading tests can serve to both make the potential users aware of hazards, (and thus avoid them) as well as prepare him or her for possible attacks.

In an article from The Christian Science Monitor by Deborah Meir, a coordinator of open education in New York City, entitled: "A Teacher's View: Reading Tests Hurt the Disadvantaged" (December 16, 1974) listed

seven reasons why standardized reading tests are limiting.

> Such tests are, by their very nature, biased in favor of children whose home and neighborhood use "standard English" and who are thus familiar and comfortable with this particular pronunciation, vocabulary, and syntax.
>
> They are biased in favor of children who have had a specific type of middle-class experience of life -- in terms of vocabulary, style of speech, knowledge of life, associations and values.
>
> They are biased in favor of children with fairly conventional, conforming, and uncreative thinking patterns -- even within that middle-class culture.
>
> They are biased in favor of children who mature early in terms of verbal skills, and thus penalize later maturers.
>
> They are biased in favor of those who think and work quickly, thus penalizing children whose thinking and working processes are more cautious or careful.
>
> They are biased in favor of children who are emotionally and socially secure under competitive and judgmental pressures. And finally:
>
> They are biased in favor of children who have been subjected to certain methods of teaching reading and certain related theories about learning to read. They thus penalize children who are learnin different ways. As a result, they tend to confirm their own biases and to lock schools into specific teaching strategies that are increasingly under question by linguists, learning theorists, teachers, and parents.

3 -- Describe the most important characteristics of the group to be tested.

This rather obvious step is a great aid to test selection. What age and grade levels are to be tested? Can one

expect the students to score above, at, or below grade level? Is English the first language for all students? Answering these questions can help the user select the correct forms of the tests to best meet his of her needs.

4 -- Understand the educationsl objectives of the program or project.

Clearly stated objectives are invaluable aids to test selection, and tests should naturally be keyed to the purposes of the program. This "natural" matching does not always occur, however, for a number of reasons. Program goal statements may not exist; goals within the program may differ so widely that clear, specific objectives are not available to the person who will be selecting the tests; and, standardized instruments may not exist to measure the most important goals.

If no goal statements exist, program decision makers should be interviewed regarding goals. In this way, both statements and inferences can be made as to goal intentions. The use of various taxonomies of educational objectives may then be useful in helping to categorize stated or inferred goals. Some sources focusing on this process include:

Taxonomy of Educational Objectives, Handbook I: Cognitive Domain, Bloom, B. S. (Ed.) Englehart, M.D., Furst, E. J. Hill, W. H. and Krathwohl, D.R., David McKay Company, Inc., New York, 1956.

Taxonomy of Educational Objectives, Handbook II: Affective Domain, Krathwohl, D.R., Bloom B.S. and Masia, B.B., David McKay Company, Inc., New York, 1964.

A Taxonomy of the Psychomotor Domain, Harrow, A.J., David McKay Company, Inc., New York, 1972.

Writing Behavioral Objectives, McAshan, H.H., Harper and Row, New York, 1970.

Goal Analysis, Mager, R.F., Fearon Publishers/Lear Siegler, Inc., Belmont, CA. 1972.

Preparing Instructional Objectives, Mager, R.F. Fearon Publishers, Palo Alto, California, 1962

Once the objectives of a course or program are clear, then one can consider which, if any, of the four major types of standardized instruments are applicable: intelligence and aptitude tests; achievement tests; personality measures; attitude scales. Or one can identify or develop criterion-referenced tests which may relate more directly to program goals.

5 -- Estimate the type and amount of resources available for testing.

Money is, of course, a prime consideration. Are sufficient funds available for the entire testing sequence? Are funds available for testing at a state or federal level?

Time is another resource factor. If tests are used to determine the effectiveness of a program, both pre-test and post-test scores should be attained, and this will require teacher time for administration and training for administration. Are the teachers willing to provide the time? (Resistance to

standardized testing often occurs because of the inappropriateness of some tests that teachers have been forced to use.) Not only are teachers or administrators willing to test, are they able to? (Some personality tests, for example, require trained administrators.) Certain intelligence tests should similarly be given by persons trained for the task.

It is also important to consider what resources are required of the local school. Some standardized testing programs include data analysis and interpretation, others do not. Thus the availablility of a computer might be an issue.

6 -- Search for Appropriate Tests.
This search may be carried out by a project person or by a hired expert. In either case, the basic question is: under the present conditions (given the goals, the population, and the resources available), what test can most effectively be used to make better decisions?

The search involves two aspects: determining both what tests seem appropriate and what the reviewers have said about them. This search can sometimes be made easier by consulting with persons in similar projects who have employed standardized tests.

7 -- Select, review, and modify the test.
Decisions concerning choices among alternative tests, considering all appropriate factors, are made during this

process by the prospective user. Then the test should be reviewed, item by item, by both the decision-maker as well as others who will be using it in order to formulate the basis for final approval or rejection.

At the time of review, the test may also be modified to suit the local user's needs. For instance, if a test of general abilities includes map-reading and work study skills, and these are not covered in the curriculum, these subtests can be eliminated. When this type of test modification occurs, it is important not to use national norms, because the testing situation is no longer the one originally used to standardize the test. In this case, local norms can be prepared which can serve as a basis of comparison within the community.

In addition to making the final decision, such as deciding not to use the test because of the language and vocabulary, or individualizing the test by eliminating certain subtests, this time period can be used to set expectations. Here various persons may make estimations on individual or group performance. Many types of comparisons can be made using standardized tests, including:

Comparison of schools within districts

Comparison of classes within schools

Comparison of students in a certain program with those not in the program

Comparisons of individual students with their classmates

Comparisons of beginning-of-year with end-of-year scores.

Comparison with previous results and original comparison of subtests with composite score to show generally weak areas.

8 -- Administer the Instrument(s)

Standardize tests require standardized administration procedures. To as large an extent as possible, tests should be given at the same time of day and during the same calendar period. This latter factor is particularly significant when using nationally normed tests. It would be expected, for example, that students who are given a fall test in December will score higher than those given at either time; the difficulty lies in comparing the results.

Teachers often require training to administer tests. Some teachers who do not follow testing procedures rigorously, especially in the lower grades, will more readily help students than other teachers who follow testing guidelines strictly. Some kind of compromise is needed in this situation.

An administration-related decision is who is to be measured: Are all students to be tested or only a portion or sample of them? Another issue is expertise of the administrators: are persons giving the tests (particularly personality tests) qualified to do this?

9 -- Interpret Results, Make Decisions

Much criticism of standardized testing is focused here, i.e., the point at which persons use the tests to make decisions. It is important to note that tests, like other evaluation products and processes, are tools -- aids -- to decision-making. Tests do not make decisions, persons make decisions based on test results, grades, direct observation, and a number of less rational reasons (like mood or the time of day).

Test publishers often cite the fallability of their instruments, but their cautions are not always followed. Standardized tests seem to be more useful in making decisions about groups because error of measurement tend to be smaller for groups than for individuals. This is particularly true when using gain scores as a measure of progress because individual gain scores are notoriously unreliable due to the variety of reasons notes earlier. In order to decrease the likelihood of abuse, test publishers often give converted rather than raw scores and present these converted scores as a range of scores rather than a specific point.

Because tests are known to have measurement errors and because testing situations can be expected to vary, even when guarded against, test users should be guided by large differences in scores between individuals or groups when making decisions. If all testees come out at approximately

the same age or grade level on a test, then other procedures should be used for making judgments, e.g., direct observation or teachers' opinions.

What decisions might be made for test results? Of the many possibilities, a few are:

Retest. Beware of making decisions on the basis of one test.

Compare test performance with other individuals.

Modify curriculum. Expand certain areas, cut back on others, include new topic areas.

Communicate the results of the testing to the parents, teachers and local community.

REFERENCES

A. Inventories and Reviews

The Seventh Mental Measurements Yearbook, O.K. Buros, 1972, The Gryphon Press, 220 Montgomery Street, Highland Park, New Jersey, 08904.

The psychometrician's bible: 1,219 tests are reviewed in this fairly expensive but readily available reference. If only one source is consulted before selecting a test, this should be it.

Tests in Print II, O.K. Buros, The Gryphon Press, 220 Montgomery Street, Highland Park, New Jersey, 08904.

An excellent bibliography of tests in and out of printing.

Improving Educational Assessment and an Inventory of Measures of Affective Behavior, Walcott H. Beatty (Ed.); Inventory prepared by Donald J. Dowd and Sarah West, 1969, Association for Supervision and Curriculum Development, National Education Association, 1201 - 16th Street, N.W., Washington, D. C., 20036.

Within categories including attitudes, creativity, interaction, motivation, personality, readiness, and self-concept, 122 instruments are listed and described. Although this can serve as a source of valuable instruments, it probably should not be the only one used for selection of tests because of its special emphasis.

The Self-Concept: A Critical Survey of Pertinent Research Literature, Ruth C. Wylie, 1961, University of Nebraska Press, 901 North 17th Street, Lincoln, Nebraska, 68058.

This volume may contain more than the Title IV evaluator needs for the selection of a suitable self-concept test; however, it is highly regarded in the field and contains over 900 references.

Measuring Human Behavior, Dale G. Lake, Matthew B. Miles, and Ralph B. Earle, Jr. (Eds.), Teachers College Press, Teachers College, Columbia University, New York, New York.

This book reviews 84 instruments and describes sources where other instrument reviews can be found. While cognitive behavior (such as reading or performing mathematical operations) is certainly "human behavior," this behavior and the tests that measure it usually focus on psychological dimensions; the instruments in this volume, however, are targeted primarily toward social interaction.

CSE Elementary School Test Evaluations, Center for the Study of Evaluation, Graduate School of Education, Copyright 1970, University of California at Los Angeles, Los Angeles California, 90024.

This volume lists 145 goals of elementary education and rates 204 tests which measure these goals in terms of validity, appropriateness to examinees, administrative usability and technical excellence. The tests cover a wide variety of skills and attitudes which might be found in Title IV projects.

B. Descriptive Books

Handbook of Research Design and Social Measurement, (2nd Ed.), Deldiert C. Miller, 1970, David McKay Company, Inc., 750 Third Avenue, New York, New York, 10017.

This volume covers the field of research design in some completeness and focuses on 35 selected instruments, mostly for sociological research. These instruments are not typically for classroom use but might be used to help judge the effectiveness of social impact or community involvement portions of Title IV projects.

Handbook on Formative and Summative Evaluation of Student Learning, B. Bloom, et.al, 1971 McGraw Hill, New York.

Described by the authors as a book about evaluating students learning, use of the evaluation procedures described should assist teachers in improving student learning. Its many sections include a unique portion presenting models and techniques for constructing valid evaluation instruments for different types of objectives on Bloom's taxonomy.

Standardized Tests in Education, W.A. Mehrens, and I.J. Lehman, 1969, Holt, Rinehart and Winston, New York.

This book, describing both cognitive and affective tests is designed for classroom teachers, school administrators and school counselors. Planning for a school-wide testing program is discussed in the final chapter.

Testing (Guidance Monograph Series, III). S. C. Stone and B. Shertzer (Eds.), 1968, Houghton Mifflin, Boston.

Project personnel might find the series of nine paperbacks within this set a basic resource for setting up a testing program. The most significant testing issues are dealt with, ranging from the history of the testing movement to the use of automated data processing.

The Essentials of Psychological Testing, Lee J. Cronbach, Frederick G. Brown, The Dryden Press, Inc., Hinsdale, Illinois.

The overview of the field of psychological testing is a good introductory-intermediate text for persons interested in the "whys" of psychological testing.

Measurement and Evaluation in Psychology and Education, R.L. Thorndike and E. Hagen Wiley, New York, New York, 1969.

Standards for Educational and Psychological Tests, American Psychological Association, 1200 17th St. N.W., Washington, D.C. 20036, 1974.

As its name suggests, this pamphlet presents specific information on a variety of aspects of test selection and use. It is a "must" for the serious test user.

C. Major Test Publishers

Educational Testing Service
Princeton, New Jersey 08540
Write to: Office of Information Service

Harcourt-Brace-Jovanovich, Inc.
757 Third Avenue
New York, New York 10017

Houghton, Mifflin Company
110 Tremont Street
Boston, Massachusetts 02107

The Psychological Corporation
304 East 45th Street
New York, New York 10017

Science Research Associates, Inc.
259 East Erie Street
Chicago, Illinois 60611

CTB/McGraw Hill
Del Monte Research Park
Monterey, California 93940

CURRICULUM EVALUATION

CURRICULUM EVALUATION

The issues concerning development and evaluation of curriculum programs sponsored by the Indian Education Act are in many ways similar to those created by the development and evaluation of any curriculum. These issues include what the curriculum is intended to accomplish, how it is to be implemented, and who is to do the implementation.

In local schools with a small percentage of Indian students, the probability of the project significantly affecting the school population as a whole might by negligible. On the other hand, in a smaller public school system in which the Indian population constitutes a larger percentage of the whole school system, a curriculum project designed to better meet the needs or interests of Indian children may have a profound impact. Yet the evaluation process could be the same.

Most curriculum efforts focus on new Indian-related programs like those seeking to teach cultural awareness or cultural heritage. Evaluation of these curricular endeavors must occur in relation to the

particular goals of individual projects, and hopefully, most of the substantive content will be more than simply addenda to the existing social studies package.

In order to evaluate existing curricula, or to evaluate new curriculum efforts, it will help to understand how curriculum comes into being. The essence of curriculum development, especially for projects involving Native Americans or other minority groups, focuses on what is to be learned. Our public school curriculum has emerged from the American middle class, and, to a degree, this curriculum reflects the social skills needed for survival in a culture dominated by middle-class values. These dominant values have, to some degree, emerged from perceptions about what constitutes human nature as well as issues relating to life fulfillment within a democratic society. However, what is assumed as knowledge, stemming from these dominant values, may not be appropriate for American Indian students, and Indian education curriculum projects may, therefore, be expected to diverge from traditional curricula.

Determining how a school curriculum relates to a general education program is sterile unless changes in classroom practices emerge. How well this curriculum is taught must therefore be investigated. It is thus useful to think of curriculum evaluation for a

given project as both the evaluation of the curriculum and of the instructional system. The quality and utility of the evaluation, as far as decisions which improve further activities are concerned, are then only as strong as the instructional leadership supporting it. An excellent treatment of the role of values in educational evaluations appears in Apple, et al., Educational Evaluation: Analysis and Responsibility (1974).

Unless the curriculum is appropriate for the educational objectives of the population to which it is directed, the program will be sterile also. The appropriateness of any curriculum therefore becomes as important a consideration for evaluation as the question of how well the program is implemented. As initial steps in curriculum development and/or planning begin with concerns regarding how the school transmits knowledge and culture, how children learn, and, how culture influences schooling, so curriculum evaluation must start with descriptions of how curriculum development occurred and of the appropriateness of the methods used.

The process of planning curriculum begins with formulated and specific goals, set either from theory or needs assessment. From these goals, specific objectives for what students are to learn then follow.

Diagnosing student learning problems and limitations is a critical component both in determining which learning activities to present and in what order, as well as in evaluating how difficult the curriculum is. Curriculum evaluation must follow a similar pattern.

The organization and selection of curriculum activities will help clarify the process of assessing curriculum outcomes. For example, criteria for bilingual education includes: learning conceptual skills in both languages, fluency in both languages, and test-taking ability in the second language (that of the dominant culture). Therefore, when an Indian bilingual project is to be evaluated, these predetermined criteria for the success of the project provide the parameters within which techniques for securing evidence and the interpretation of the data become meaningful and valuable.

Questions concerning objectives of the program are only one part of a comprehensive evaluation. Other considerations focus on ways in which students achieve those objectives, how thoroughly the curriculum is learned, and the usefulness of the curriculum. Further, information concerning other aspects of the educational system, such as the instructional program and the guidance program, will also be part of a comprehensive curriculum evaluative effort; e.g., might the goals of

a history class be explained to parents in native rather than dominant language? Do counselors speak both languages? Are classroom techniques in nineteenth-century American history taught from the vantage points of both cultures? The range of questions possible in curriculum evaluation is discussed in "If Everyone Agrees . . ." and "If Nobody Agrees . . ." (House, 1973).

Efforts in creating a new curriculum project are often scattered or watered down when there is little attempt to integrate the innovations of the new project into the overall curriculum. For example, creating a new guidance program for Indian students without changing the guidance program in the traditional program for other minorities diminishes the impact of the effort by not decreasing (and, in fact, possibly heightening) feeling of inferiority. How the curriculum is set into the school and how various aspects of the system as a whole support the curriculum become crucial aspects of curriculum innovation development and, ultimiately, evaluation. Issues involved in curriculum evaluation are treated in Wittrock and Wiley (1970) and include several sociological and psychological, as well as educational concerns.

A final thought in considering how to begin to evaluate curriculum programs concerns the dual issues

of leadership in a new program and collective teamwork. Making it possible for all new participants in the new program to feel responsible and useful is a function of the program leaders. These feelings may, in turn, be the key to the acquisition of unbiased data. The degree to which a new or special project becomes a part of the school system is related to the degree of involvement on the part of other teachers, educators, or pupils.

Basic Conceptualization

Many of the projects of Indian Education Act programs focus on developing and teaching a special or specialized curriculum. A number of these can be grouped under a general category called cultural heritage projects which typically maintain one of two directions: they either attempt to bring together groups of indivduals with different ethnic backgrounds or they try to teach a specific group about its own history and traditions to enhance self-concept, enrich perception, and enlarge life possibilities. Curricula in these projects are often structured around ethnic art, language, or music, and history.

Special Projects

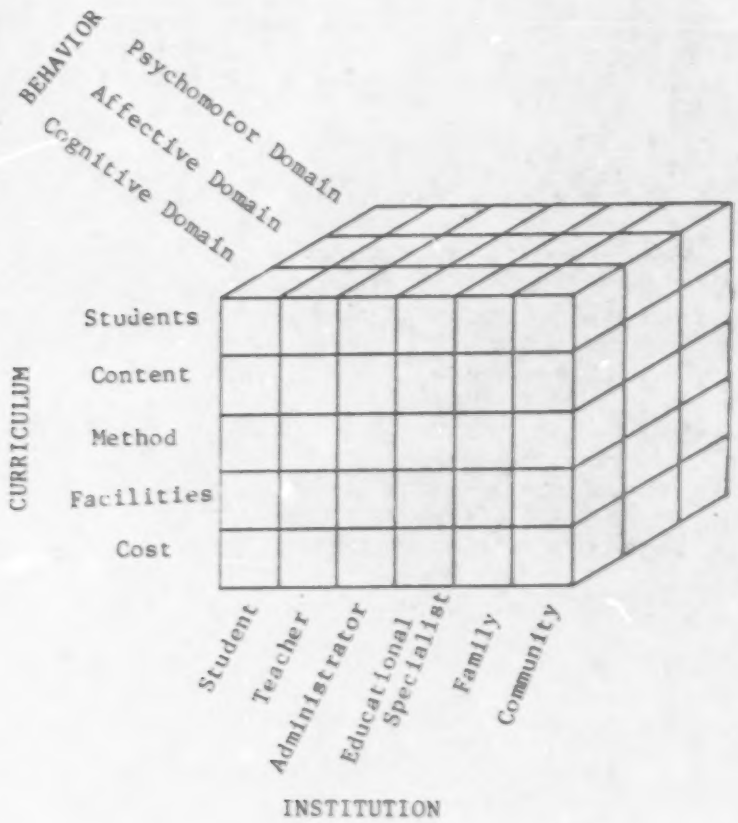
Cultural heritage programs usually are offered outside the regular school curriculum. Whereas it has been the practice to assess regular school curricula

by means of standardized achievements tests and curriculum implementation (process) studies within the administrative structure of the district, cultural heritage projects or specialized curriculum projects have been assessed by individually tailored evaluations. In fact, the evaluation of these special projects has provided the structural basis for many evaluation models in current use, and can provide additional frameworks for curriculum evaluation, regardless of particular type.

The several models that have been developed through evaluating projects of this type represent the different methods professionals have employed in evaluating special projects. Perhaps the most noted of these models are: the CIPP model (Stufflebeam et al., 1971), the EPIC model (Worthen and Sanders, 1973) and the Discrepancy model (Provus, 1971). In setting up criteria for measurement of special curriculum programs, some of the most basic creative work may be done, since there are few if any standardized measurements for the objectives of these programs and there is an absence of similarity or comparability among projects -- even those within the same curriculum area.

Generally, these models present sound conceptual frameworks through which the techniques associated with evaluation can be synthesized and applied with consistency.

For example, the EPIC model provides a two-dimensional display of a three-dimensional configuration that leads one to observe and define interaction of three dimensions in order to identify the array of evaluative concerns or questions.



The CIPP model presents evaluative questions categorized according to four areas of program operation: Context, Input, Process and Product. The Discrepancy model couches evaluative questions in terms of gaps or discrepancies between what was intended and what is observed. For example, if the designer of the project intended that students investigate standards of literary criticism and,

in conducting the class, the teacher presented a formulated set of standards, a discrepancy between intention and observation would be evident.

None of these models provides a panacea or a "right and only" way to conduct an evaluation, and none of the models has been perfected or standardized to the point where any one can be applied readily to situations other than those for which they are tailored. Applying one of these models to another situation requires careful study and expertise. Text materials are available to guide in the selections and application of models. (Worthen and Sanders, 1972, have provided an especially useful matrix which can be used for quick comparison of the various models).

Basic Procedures

Whether the program evaluation models abstracted in Worthen and Sanders or an eclectic approach is chosen, a discussion of several major common elements of an evaluation (e.g. relationship to decision-making, relationship to objectives, role of evaluator, key emphasis, purpose, implications for design, criteria for judging evaluation, constructs proposed, limitations and contributions) should be useful here, not only to get a feeling of what they are, but also to get a feeling for the actions

and movements of an evaluator at work.

It is important that an evaluator begin the work day with the curriculum design staff. Although this practice is not followed in many cases where insufficient evaluation occurs, this daily involvement, especially in the initial phases of the project, is invaluable. The evaluator or program design staff should have available needs assessment information or should plan to gather information prior to deciding what the curriculum is to accomplish. Too often projects are created to get available funds rather than to meet needs. Ideally, they should be designed to meet needs, and the needs assessment information should become the basis for project objectives and evaluation standards.

In designing an evaluation, the evaluator undergoes a process for acquiring knowledge essential to complete his work. Care should be taken to digest thoroughly all materials needed to formulate the approach prior to the attempt to design the evaluation. In addition, efforts should be made to consult often with experts within the curriculum area being evaluated. Input from experts familiar with the subject matter areas of the program being evaluated can be extremely useful in making selections of scales and tests and in anticipating results. The evaluator, by interview, can learn from the expert what he is

looking at in making his judgment so that that knowledge can be used in constructing a scale to assess workmanship.

The following presentation is an attempt to describe both the evaluative process and a typical evaluator at work. Alternate methodologies will be referenced and in some cases briefly described. However, these descriptions should not be construed as a guide to the application of the methodologies, but only as suggested reference sources to be consulted prior to attempted application of any one methodology.

Project Description

In both designing and reporting an evaluation, description is one of the most important activities of the evaluator. In order to know what to look for, to know to whom to report information, and to know when and how to report the information, the evaluator must become familiar with all facets of the project. He/she must know what different people associated with the project want the project to do, to whom, when, and how. Hence, description of the project is essential in helping to identify: components to be observed and tested, expectations held for the project, the basis for judgment of worth of the project, and what is expected of the evaluation.

By an analysis and detailed specification of the objectives of the project, the evaluator can pinpoint criteria for successfully meeting project objectives and

make the first approximations toward development of acceptable observation instruments. On this basis, a start can be made to determine manpower requirements (how many data gatherers are needed), the level of knowledge sophistication (how many modes of analysis may be required), and written output requirements (how many reports must be submitted at what times within the project period).

A further descriptive obligation of the evaluator is to present the project in such a manner that someone reading the evaluation report can determine what activities occurred during the project and what the project was about. This description would address itself to such elements as:

- Age and characteristics of target population
- Number of participants
- Procedures to be implemented
- Materials to be used
- Time period (s) for various stages of the process
- Measurement techniques and analytic models employed
- Results.*

To a degree, it can be argued that the evaluator's prepared description of the project should be of such quality that from it someone can replicate the project. This

*(Hawkridge, D. Writing an Evaluation Report, Palo Alto, 1971.)

description can be made easier if the evaluator does not have a lengthy history to delve into (the gathering of historical information in an ex post facto manner has been found to be both frustrating and time-consuming). However, this is not to downgrade the importance of contextual information within the project. The evaluation should reach back as far as possible to the start of the planning stages of the project or to the needs assessment study. For example, needs assessment studies in curriculum often involve inferences from achievement tests or require data from proposed courses.

Descriptive Methodologies

The evaluator selects one or two of the several descriptive methodologies to begin the process. He combines this task with ample conversation with decision-makers in the project to assure that he is not functioning in a vacuum. Thus, the evaluator chooses a descriptive methodology that will serve to describe the objectives, intents, or purposes of the project, such as: behavioral objectives (Mager, 1962); expressive objectives (Eisner, 1967); task delineations (Gagne, 1967), or utilization of a "fuzzy concepts" approach (Hutchinson, 1974). The evaluator will use this methodology both as a framework for establishing the dimensions and scope of his work and as an integral part of the reporting design.

Charting Techniques

The evaluator also chooses a second form for describing the project, usually one or more types of charts that display different relationships within a project. The evaluator elects to use either a list of decision-makers or an organizational chart to identify the components of the project. He will often choose a flow chart or a case study approach to try to describe how the project works or how people gain access to, participate in, and leave a project. Some evaluators prefer an activity list. Also, the evaluator will choose either a GANTT or PERT charting technique, relating different people to materials, facilities, and functions at different times to describe the project's chronological progression. (Excellent treatments of charting techniques can be found in Cook, "Program Evaluation and Review Techniques, Applications in Education," 1966; and in Borich, Evaluating Educational Programs and Products 1974, which includes a sample activities chart of a curriculum study.)

Determining Information Requirements

After the evaluator has garnered adequate descriptions of the project, he develops a plan of work. As the initial step, he selects from his charting descriptions a variety

of people concerned with the project evaluation and confers with them about evaluative information. This step is referred to as audience indentification and is usually the evaluator's first attempt to tackle the problem of the various levels inherent in decision-making (Stufflebeam, et al., 1971). In the case of curriculum, it may include parents, teachers, school administrators and pupils. These interviews may or may not be structured or even recorded. However, records will be useful later for design reference. From these interviews, the evaluator gains a grasp of information requirements in terms of both content and time.

Evaluative Questions

This knowledge is then translated into a list of evaluative questions. An illustration of potential questions can be designated by phases of program operation (Worthen, 1973):

- What were the objectives of the Native American heritage program?
- How many participated?
- What personnel factors were associated with the accomplishment of these objectives?
- What was the quality of the instruments and materials used in pursuing the objectives of the program?

Sometimes this list becomes the basis for the actual evaluation, but, most often, it is too exhaustive to be practical in terms of time, resources, and other constraints. The next step is to reduce the list of specific information needs in order to set priorities and limits to the evaluative task by asking decision-makers to identify priorities and determine which evaluation queries are most critical to answer.

Measurement Devices

Once the questions have been decided upon, the next task is to establish criteria and observation techniques needed to answer them. Usually the evaluator returns to the description of the objectives and identifies a range of possible measurements for each objective. If the curriculum objectives are written to describe audience, behavior, condition, degree, or outcome symptoms, measurements are easier to identify. Sometimes a table of relationships between information requirements and potential criterion measurements is useful (Stufflebeam, et al., 1971). This range of measurement may include standardized tests, previously developed scales, plans to develop a scale or test, record-keeping system, observation and interview schedules, or self-reporting

procedures. The evaluator then returns to his audiences and/or decision-makers and solicits their preference of measurement devices to be employed. In the case of curriculum, this is important so that correct interpretation of the objectives can be assured and so that specific tests of content can be planned for the assessment of achievement of the content in the curriculum. Emerging from this process will be the criteria for selection which may include the quality of the measurement (reliability, validity, etc.), the cost and difficulty of applying the measure, and the "face validity" of the measurement to the decision-maker.

A similar procedure is involved in converting the descriptive information gathered through the charting techniques into a measurement tool through the use of taxonomies and coding manuals. Sometimes it is useful to utilize two different charts. One chart could depict the organization from which levels of decision-makers can be identified (often these levels are associated with the level of abstraction of evaluation questions and the levels of data aggregation, such as classification by class, school or district). The other chart could depict either relationships of components or functions and/or activities of people by time.

Input Specifications

In the course of establishing measurement controls, the evaluator is invariably confronted by the question of evaluative dimensions. Input specifications will be required for such diverse categories of data as staff size, space, materials, time, personal interactions, staff morale, encumberances, past expenditures, cash on hand, etc. Again, a range of techniques for coping with the information must be explored if the evaluation is to be manageable within the constraints of resources available and with control over the quality of the results. The evaluator again consults the decision-makers so that the choice of measures is restricted to those favored by them. (The evaluator should be further restrictive so that measures are chosen with psychometric indices reflecting optimum quality.)

Analysis Plan

Once the question of measures has been decided upon, the evaluator is still faced with several design decisions, some of which are contingent upon the capability and availability of staff and resources. He must design an analysis plan which will summarize the data and allow him to report the results at the times required by the

decision-makers. The analysis plan should take into account the quantity and quality of the data, thrust of questions, types of measures, and the complexity of the reporting task, since the evaluator must estimate the capability of the audiences for whom the report is intended to comprehend statistical results. Several texts deal with analytic models. For the beginner, one of the best is Reading Statistics and Research by Huck, et al. (1974); for the more advanced evaluator, the Kerlinger/Pedhazur test entitled Multiple Regression in Behavioral Research (1973) is extremely useful. In some curriculum studeis, the primary analysis question simply involves subjects learning the content or changes in the curriculum directions. In other curriculum studies, comparison with control groups or competing models is the primary question.

Data Collection

In addition to the completion of the analysis plan, the evaluator must detail procedures for collecting data. These include the training and monitoring of data collectors; editing and, in some cases, coding; storing these data; control of data so that none is lost and so that poorly administered data can be identified.

In some cases, make-up tests and documentation of data must be planned. Care and supervision must also be

planned if teachers are used to collect data on curriculum. In many cases, data controls and storage can be managed adequately through an expanded set of subscripts which will be referenced later. A good checklist for data collection methods is provided by Worthen and Sanders in Educational Evaluation: Theory and Practice, 1973.

Monitoring

Suffice it to say, the evaluator should plan to monitor his/her activities, and when the evaluation represents a major effort, attempts should perhaps be made to contract for a meta-evaluation (an evaluation of the evaluation, which will also be referenced.)

Once the evaluator has detailed both the analysis and the data collection activities, he/she must go through a tooling-up phase which includes pretesting the analysis procedures or programs, training the data collection staff, piloting the data collection activities, and designing checks and audits to document the quality of the operation. The plan is then implemented, but work is not complete, even after the data have been analyzed.

Reporting

The data must now be reported so that they translate into consumable information. This reporting process must include a specific attempt to write the material in

the form most accessible to each audience, to document its translation to each audience, and to provide a final overall package not only of what was reported to each audience but to attempt to relate the results to existing archives. In doing this, the evaluator must document everything, continually reanalyze his/her role, and try out the data collected on project-associated people to provide a basis for any judgment that has to be made in midcourse. Again, Worthen and Sanders provide a good summary of accepted practice in evaluation reporting.

In attempting to describe how a typical evaluator might evaluate a project designed to teach a special or specialized curriculum, the tasks of the evaluator has been described. No attempt has been made to describe his competence, his commitment, his roles and self-conception, or his techniques. The description has included allusion to several terms which have acquired technical meaning, and has mentioned several methodologies, techniques, and activities, the specifics for which may be found in references mentioned concurrently. A dictionary of terms may be found in The Encyclopedia of Educational Evaluation (Anderson, Ball, and Murphy, 1975). A list of references keyed to the steps in the curriculum evaluation process is also provided.

REFERENCES

Anderson, S. B., et al. Encyclopedia of Educational Evaluation: Concepts and Techniques for Evaluating Education and Training Programs. San Francisco: Jossey-Bass, 1975.

Apple, M. W., et al, Eds. Educational Evaluation: Analysis and Responsibility. Berkeley: McCutchon Publishing, 1974.

Borich, Gary D., ed. Evaluating Educational Programs and Products. Educational Technology Publications, Englewood Cliffs, New Jersey 1974.

Cook, D. L. "Program Evaluation and Review Techniques, Applications in Education." U. S. Office of Education Cooperative Research Monograph, No. 17, OE-12014. Washington, D. C.; U.S.O.E., 1966.

Eisner, E. W., Educational Objectives: Help or Hindrance? The School Review, 1967, 75, pp. 250-260.

Gagne, R. M. "Instructional Variables and Learning Outcomes." Paper presented at the UCLA Symposium on Problems in the Evaluation of Instruction, Los Angeles, California, December, 1967.

Hawkridge, D. Writing an Evaluation Report. American Institute of Research, Palo Alto, California, 1971.

House, Ernest R., Ed. School Evaluation: The Politics and Process. McCutchon Publishing, Berkeley, 1973.

Huck, S. W., W. H. Cormier, and W. G. Bounds. Reading Statistics and Research. Harper and Row, New York, 1973.

Hutchinson, Thomas, "The Operationalization of Fuzzy Concepts," Paper presented to AERA Division H, Chicago, 1974.

Kerlinger, Fred N. and E. J. Pedhazur. Multiple Regression in Behavioral Research. Holt, Rinehart & Winston, 1973

Mager, R. F. Preparing Instructional Objectives. Fearon Press, Palo Alto, California 1962.

Provus, Malcolm. Discrepancy Evaluation: For Educational Program Improvement and Assessment. McCutchon Publishing, 1971.

Stufflebeam, Daniel, et al. Educational Evaluation and Decision Making. F. E. Peacock, Itasca, Illinois, 1971.

Wittrock, M. C. and D. E. Wiley. The Evaluation of Instruction: Issues and Problems. Holt, Rinehart & Winston, New York, 1970.

Worthen, Baine R. and James R. Sanders. Educational Evaluation: Theory and Practice. Charles A. Jones Publishing, Worthington, Ohio, 1973.

PERSONNEL EVALUATION
AND
IN-SERVICE TRAINING

PERSONNEL EVALUATION AND IN-SERVICE TRAINING

This paper deals with elements of an overall evaluation plan that relate specifically to measuring the impact and effectiveness of in-service training under the Indian Education Act. Though specific focus is evaluation of in-service teacher and administrator training, methods and techniques apply generally to any evaluation of teachers, administrators, or curriculum.

The phrase "elements of an overall evaluation plan" is emphasized because overlap and duplication in the discussion of techniques and methodologies described in other papers will occur: regardless of the specific target of evaluation, variations on the same techniques--interviews, observation, standardized tests, written forms, etc.--will appear and reappear.

It is hoped that these papers--and especially of this presentation--will enable an evaluator to make multiple use of information gathered for a single purpose. By viewing in-service training as a factor in measuring the impact of reading or special curriculum programs, for example, the creative evaluator can save both time and resources by designing a data collection system that can, simultaneously, serve the evalutive needs of several project components.

In this context, the evaluation of in-service training of teachers and administrators as part of an overall evaluation plan will be discussed.

1. Assessment of performances in relation to training goals.

2. Means of assessing those performances.

3. Standards of comparison between intent or objective and performance.

4. Reporting and other use of evaluative information.

1. Assessing Performance of Teachers

Most in-service training programs will include some or all of the following types of training goals:

-- To improve teaching skills.

-- To modify or change attitudes.

-- To change the teacher-pupil interactive process.

-- To provide teachers with curriculum materials and ability to evaluate curriculum materials and resources.

-- To improve teachers' ability to assess pupil achievement and to diagnose individual problems.

-- To develop interpersonal communication skills.

-- To sensitize teachers to the special needs of Indian students.

The goals of in-service teacher training should help to identify and classify teacher behaviors within which the parameters for evaluation can be developed.

An underlying assumption of the evaluation of in-service training is that teacher behavior affects pupil learning. In an evaluation design, in contrast to a research design, however, no attempt can be made to establish a causal relationship, and no attempt should be made to use pupil learning as a criterion of the effectiveness of the teacher training. Performance to be assessed will be limited to behavior of teachers.

Assessing the impact or effectiveness of the training program in furthering the training goals is different than evaluation of other program components. Although many of the goals of training are subordinate to other goals (i.e., teachers are taught to teach better so students will learn more), it is not possible in many cases to assess results achieved through training in terms of student attitudes, behavior, and performance at some future date. A more practical approach is to base evaluation on the objectives of the training program as they relate to teachers and assume that these objectives were set as part of a valid design to achieve the desired long-range results with students. In the broad sense, the purpose of in-service training is to bring about a desired modification of <u>teacher behavior</u>, whether through increased interpersonal sensitivity or orientation to new developments in a curriculum field. The two essential elements in a training program designed

to change teacher behavior are: (1) the subject matter of the training, and (2) the process involved in conveying the subject matter. An evaluation of the program must, therefore, consider these elements.

Improvement of Teaching Skills

Determining the appropriateness and priority of skills training is dependent on needs in various projects, and the specific teacher behaviors requiring evaluation will depend on which skills have been given priority. Evaluation of skills training answers questions such as: "Can the teacher use this skill?" "With what frequency does the teacher use the skill?" "Does the teacher allow for traditional cultural response patterns in applying the skill?" "Does the teacher use the skill in its intended way?"

Attitudes

Attitudes of teachers toward pupils, community, parents, and diverse cultural patterns comprise important performance dimensions of an in-service training program. Measurement of attitudes therefore reveal the individual's beliefs, feelings, and action tendencies with respect to an object (or person). Specimens of behaviors must be carefully selected but will be dependent upon the type of attitudes underlying teacher behavior toward which attempts at change

are directed. Again, the performance to be assessed is a teacher behavior, either observed or reported.

Interactions

Interactions of teachers and pupils resulting from a teaching-learning situation comprise another behavioral dimension to be assessed in an in-service training program. Questions frequently asked in evaluating teacher-pupil interactions include: "How directive is the teacher?" "How much does the teacher allow pupil ideas into the classrooms?" "What percentage of time does this teacher talk? Listen?" "Is there a specific pattern of interaction with respect to makeup of the classroom, e.g., Indian/non-Indian?" Some observation instruments have been developed to measure these kinds of interactions.

New Teaching Materials

Use of new teaching materials is the most common mode of conducting an in-service teacher training session. The types of performance to be assessed when evaluating how this material is used will again be teacher behavior. Questions include: "Does the teacher use the new material?" "Does the teacher understand the new curriculum?" "Has the new material been properly evaluated to identify discrepancies and errors in fact?" "Is bias discernible in the material?"

"Are intentional or unintentional statements present which may cause emotional damage to Indian children?"

Assessment of Pupil Achievement

This is to be a priority task and performance is somewhat different from skills training or attitude measurement. Questions asked in this area may include: "Can the teacher relate tests to teaching objectives?" "How well can the teacher interpret standardized tests?" "Can the teacher interpret test data to pupils?" "To colleagues? to parents?" "Do the teacher-made tests assess essential achievements of the pupils?" "Can the teacher communicate effectively with parents? with administrators? with students, both Indian and non-Indian? with other school related persons?"

Sensitizing teachers to the special needs of Indian students is an implicit objective of the Indian Education Act, and such a goal would most likely be served through training in awareness of Indian culture and heritage. The performance to be assessed will again be teacher behavior demonstrating sensitivity to cultural factors.

2. Means of Assessing Performance of the Trainee

The type of trainee behavior to be assessed will normally dictate the most appropriate means for assessment. Potentially, most behaviors can be assessed by several methods. Individual judgment based on such factors as

economy, objectivity, and documentation should be exercised to select the most appropriate assessment means. Multiple assessment is strongly recommended whenever possible.

The following table is presented to help identify the most appropriate means for assessing any given behavior. A variety of means are listed across the top, with categories of performance down the left margin.

East assessment technique relates to specific teacher performance or behavior categories and requires three steps: (1) recording of behavior; (2) the classification of the behavior; and (3) analysis of this data to provide information for decision-making.

ASSESSMENT MEANS

PERFORMANCE CATEGORIES	Videotape	Audiotape	Direct Observation	Pencil & Paper Tests	Colleague Rating	Self-rating	Pupil Rating	Parent and or Parent Council Reporting	Survey or Interview	Work Sample
Skills	X	X	X							
Attitudes	X			X	X		X	X	X	
Interactions	X	X	X				X	X		
New Curricula	X	X	X	X	X	X				X
Pupil Assessment				X						X
Communication Skills			X		X	X		X	X	
Cultural Sensitivity			X	X	X		X	X		

Recording Behavior (Taping)

The most objective method of collecting data about teaching is to record the actual performance with either a video camera or audiotape.

Two factors must be considered in using these techniques: (1) intrusion of any device in a classroom may in itself cause a change in the classroom environment, and (2) recording of several hours of classroom activity will necessitate a multiple of those hours in classifying the behavior.

The first problem can be only partially resolved. If using videotape, the camera must be placed in the room for several hours, preferably some time each day, before it is used to record the segment of behavior desired, since awareness of surroundings diminishes with familiarity. Also, it is advantageous to allow the teacher and students to videotape and replay short segments of activity so that the camera becomes a familiar object.

The major drawback to audiotaping, although the equipment is less obtrusive, is in identifying the speaker. One way to overcome this second problem may be to introduce the equipment through in-service training as a lesson-developing or learning set. The video camera should thus

be used only to record this aspect of a lesson or series of lessons. This process might be termed selective recording, that is, recording of the particular portion of a classroom day wherein the specified behavior to be observed is most likely to occur.

However, it is unlikely that selective recording would be valuable for observing, for example, a teacher's skill in questioning. In this case, employment of a sampling technique would be more suitable for the purpose.

Autiotaping facilitates data collection, sampling, and assessment of teacher-questioning skills. A variation on the technique is to use a remote microphone, placed around the teacher's neck, with no other obtrusive devices in the classroom. The information is recorded in another room with a timing device turning the recorder on and off at pre-established intervals.

Observation

A less objective but still valuable data-collecting method pertaining to teacher skills is direct observation. An advantage of direct observation is that it is possible to classify behavior simultaneously with recording. A disadvantage is that the behavior is not directly reproducible and therefore cannot provide feedback to benefit the teacher. When behavior can be reproduced by rewinding

or stopping a video tape to examine a classroom scene closely, the evaluator, administrator, or teacher is much more likely to reach agreement that a certain behavior did, in fact, occur. Direct observation, while not reproducible, may lend itself readily to certain frequency counts so the data is classified as it is recorded, and expense is reduced.

In using direct observation, categories of observations must be selected and defined, observers trained to recognize the behavior, and reliability of observations checked. The classification of observed data is an essential pre-step to analysis. Usually, this is a three-step process: (1) the selection of categories of behavior to be classified, (2) the definition of the categories, and (3) the examination of the data collected (videotape, audiotape, etc.) to count the frequencies of the behavior.

Because categories of behavior will be dependent upon in-service training objectives, some guides may be valuable in establishing these categories:

(a) Define the behavior in terms on action verbs.

(b) Be sure that categories do not overlap.

(c) Include sufficient categories to permit a complete description but not so many that refined observer judgment precludes inter-observer reliability.

Classification should be performed by paris of observers working independently. Inter-observer agreement must be regularly verified.

Pencil-and-Paper Tests

Pencil and paper tests refer to any of a large number of instruments for assessing teacher behavior. For example, if data is wanted on whether teachers comprehend a set of new curriculum materials, a series of tests may be administered. When assessing teacher attitudes, an attitude instrument may prove to be most economical and efficient.

Ratings

Various kinds of ratings (colleague, pupil, self) can be a valuable technique for assessing certain interpersonal behaviors when used in combination and in conjection with other behavioral considerations. If, for example, a teacher's self-rating of a behavior reveals a considerable discrepancy between it and a colleague or pupil ratings of the same behavior, this discrepancy can then be addressed and behavior of the teacher may begin to change.

Parents

Parents and/or parent councils are particularly valuable sources of information on teacher behavior in the areas of attitudes, communications, and cultural sensitivity. Reports from this sector may range from informal data through survey or interview forms to rating

scales. Emphasis should be placed on parent reporting for follow-up on evaluation of training sessions.

Community Survey

It may be desirable to contact the community with a survey instrument to assess the impact of certain training programs, particularly in communications skills and attitudes (The survey is also valuable for evaluating the in-service training process itself).

Inferences drawn from either a mailed form or an interview by telephone or in person must be cautious unless a high degree of representativeness has been achieved. Often the most frequent respondents to mail surveys are those who have strong, negative feelings. However, the alternative of one-to-one interviewing is extremely time-consuming and requires trained interviewers to obtain frank and accurate responses. If sampling is employed with either a survey or interview technique, great care must be taken to assure that the sample is valid or that the sample is, indeed, a random one.

Work Sample

A work sample examines some tangible teacher products: if pupil assessment has been a training goal, copies of teacher-made tests may be examined to guage adherence to principles taught in the training session; anecdotal records compiled by a teacher may reflect the teacher's comprehension

of new curricula.

In summary, the means of assessing behavior will always be determined by the behavior to be assessed. Considerations of economy and objectivity will also influence the decision, and obtaining multiple measurements of any behavior whenever possible remains desirable.

Evaluation of Administrator Training

The success of an Indian Education Project will require training of administrators in addition to training of teachers. Again, it must be stressed that multiple use of information gathering is essential. Further, if a project does not have an in-service training component for training of administrators, the material that follows may still be useful in terms of general evaluation of administrators or of some aspects of curriculum evaluation.

As with training of teachers, administrators will be trained to achieve certain desired objectives. The assessment of whether or not these objectives have been achieved will depend upon identifying evidence of administrator behavior aimed for by the training.

An administrator's tasks can be categorized as planning, organizing, directing, or controlling activities requiring a variety of skills. The following types of training goals, although not exhaustive, should serve as examples of those generally included in an administrative training program:

- To improve curriculum planning.
- To develop supervising skills.
- To improve the teacher evaluation process.
- To develop interpersonal skills.
- To acquire goal setting skills.
- To modify or change attitudes.
- To sensitize administrators to the special needs of Indian students.

Because some of these goals are similar or identical to in-service training program goals for teachers, assessment procedures will also be similar.

Videotaping, audiotaping, and to some extent direct observation are not useful means of assessing administrator behavior; and skill acquisition, attitude change, and cultural sensitivity will require less direct observation than with teachers. In most instances, administrative behavior can be adequately assessed through work samples, ratings, and surveys or interviews.

Prior to determining assessment means, each training goal needs to be clearly specified and questions phrased as to what behaviors are desired as outcomes of the training.

Curriculum Planning

This goal involves a complex set of process and product oriented skills. The process skills are basically interpersonal skills and will be dealt with later. The product skills will be observable in terms of the emergence of a

of a curricular plan. The administrator must obviously play a leadership role in developing school curricula. Within this area, such questions might be phrased as: Is the curriculum up-to-date? Is it appropriate for the age level? the cultural/social make-up of the school? What educational philosophy or pedagogical theory ties the curriculum together? Is the theory compatible with the goals of the project?

Supervising Skills

This area of behavior represents a special category of interpersonal skills--those dealing with the management of people resources. The administrator must be sentitive to the needs of the teacher as well as capable of meshing those needs with the goals of the school. Questions which might be asked in this area include: Do teachers know what is expected of them? Does the administrator have a clearly stated set of goals or objectives? Is the administrator approachable? What is the degree of teacher satisfaction or morale? Is there a high turnover?

One subset of supervising skills is used in the process of evaluating teachers. Administrative responsibilities that call for skills in this area would be assessed by asking such questions as: Does the administrator have a regular, systematic plan for evaluation? Is the evaluation conducted from a multiple measurement base? Does teacher

observation play a major part in the evaluation? Does the administrator hold regular conferences to discuss strengths and weaknesses with individual teachers?

Interpersonal Skills

During any given day an administrator deals with a variety of people: teachers, students, parents, other administrators, persons in other service systems. The interactive process is far more likely to affect the quality of results than the content of these meetings. An evaluator would be therefore asking questions such as: Is the administrator directive or nondirective? Does the administrator listen? Are interactions conversations or monologues? Does the administrator continually focus on the problem rather than on personalities? Is the administrator considered fair?

Goal Setting

Expectations must be mutual for maximum effectiveness of an organization. When a teacher knows what is expected and this coincides with what the administrator expects, supervision, evaluation, and most other management tasks become relatively easy. Questions dealing with an administrator's goal-setting skills might include: Are goals clearly stated? Are the expected outcomes clear or fuzzy? Do the goals represent mutually negotiated expectations?

Does the administrator demonstrate that these are "real" goals in terms of his decisions?

Means of Assessing Performance

Whether evaluating the effects of training, the performance of the administrator, or some other aspect of the program, certain principles are basic: the more direct the observation, the more useful it is; multiple observations are desirable whenever possible; assessment of behavior, either reported or observed, is the basis for evaluation; and evaluative observations should be planned for multiple uses for curricula, teachers, administrators, or whatever program uses to which they can be fruitfully applied.

Direct observation of administrators' behavior are somewhat less available than those of teachers because their activities are generally less confined, in both frequency and location. Performance assessment techniques will generally fall into the following categories:

- Paper-and-pencil reporting
- Ratings
- Audiotaping
- Work Sample

Paper-and-Pencil Reporting

Two types of paper and pencil reporting are generally

used: self-reporting and statistical or archival data. Although archival data indicating the frequency of occurrences is generally the most reliable, however, care must be taken to ensure that what is reported over a period of time is adequately defined and compatible over that time period. Self-reporting generally consists of one's perceptions of one's own accomplishments, abilities, skills, or attitudes. In considering the question of the frequency of teacher conferences or of classroom visits, the administrator should be able to provide such data with a high degree of accuracy and reliability from existing records, such as a daybook or calendar. Whether one feels that he or she has adequate abilities or skills can then be reported on a check list or questionnaire.

If one considers the administrator's attitude to be important, perhaps self-reporting is not sufficient as a means of assessment. To achieve multiple measurement, reports may be requested from teachers, other administrators, and parents in terms of observed behaviors which create perceptions of attitudes. Those instruments in the general category of paper-and-pencil reporting usually require some technical expertise to develop and should be carefully designed, tested, and validated before being used for decision-making. (This of course, does not apply to archival records).

Ratings

Rating scales comprise a special category of paper-and-pencil reporting with which multiple observations are strongly encouraged. When ratings are designed with behaviors clearly specified and administered to obtain several perceptions, they can be very useful. Self-perceptions compared with those of supervisors, colleagues, and pupils may yeld valuable results.

Audiotaping

Limited use of audiotaping may prove fruitful. Blumberg (1974) describes uses of interaction analysis involving audiotaping similar to the Flanders system for classroom observation (1970). This method of assessing behavior is particularly useful for assessing supervisor and interpersonal skills when such activities take place within the administrator's office. (A complete description of this procedure is available in Simon and Boyer, Mirrors for Behavior VIII, 1970).

Work Sample

Products of an administrator's efforts may provide useful data. If supervisory skills are being assessed, the staff handbook, for example, might be a product worth examining. Similarly, plans for staff development or in-

service training for which the administrator has been responsible might provide evidence that such skills are satisfactory. Records concerning disposition of student disciplinary cases when examined in terms of Indian vs. non-Indian students may reflect the degree of administrator sensitivity to the special problems of Indian students.

Evaluation of the In-Service Training Session

In addition to the subject matter covered in a training session, the process of training should facilitate trainee growth. In order for participants to answer questions concerning the value of the training, assessment should be made at the time training is in progress.

A good starting point is assessment of the trainer or training staff. What qualifications did they have? Were they capable of meeting participants' training needs? Did the participans perceive them as helpful? Did the trainer demonstrate biases which negated his training?

How did the participants leave the sessions? What kinds of feelings did they have? What additional needs of the trainees emerged from the sessions? A combination of survey and rating forms should be the most valuable data collecting devices for this type of assessment.

Standards of Attainment

If "need occurs when what is happening is below that which is expected," then determination of "that which is expected" is, in effect, determination of the standard of attainment desired. Comparison of performance must continually be made in conjunction with the statement of needs and a judgment as to whether the need has been fulfilled. Acceptable standards of attainment, then, are fulfillment of needs.

The need to "encourage children to express themselves in the classroom" can be translated into several objectives, one of which may be to train teachers to become less directive. Under the performance category of teacher-pupil interaction, this would result in assessment of the teacher-pupil behavior through such means as the Flanders Interaction Analysis Matrix (Flanders, 1970).

To use another example: If a need is identified in the area of teacher attitude or cultural awareness, the objective is to train teachers to accept cultural differences. Assessment of the accomplishment of that objective might include a survey of the pupils or of the parents and pupils. All of these would be tested as a relative standard for determining whether there has been some degree of sensitivity.

Another method of measuring such a performance may

include an attitude scale/self-reporting instrument. Since standards are inextricably linked to measurement, failure to measure an appropriate performance may be caused by factors other than lack of performance, i.e., the assessment instrument may be insensitive to the behavior being assessed; or, the measurement may be an attempt to assess the failure to meet a standard when, in fact, the standard has already been met. For example, a survey may seek to determine how many complaints have been made about a teacher's insensitivity to cultural characteristics different than his or her own. Failure to find complaints may indicate that there are no complaints, or that the instrument failed to detect the complaints, that the wrong standards are being compared. For example, the evaluator may be looking at the degree of teacher directiveness in teacher-pupil interaction to account for negative reaction to work incentives, when, in fact, the teacher-lack actually lies in the area of sensitivity to a culture in which the pupils are reluctant to compete against their peers.

Reasonable and appropriate standards are thus essential to success in using the information collected for decision-making.

Reporting and Use of Information

Both formative and summative evaluation are designed

to serve decision-makers. Evaluative information obtained from the in-service training component will be essential for both formative purposes and summative ways.

The importance of sharing evaluation results cannot be overemphasized; immediate feedback is essential. But evaluation must serve a wide variety of audiences, and information appropriately compiled will be needed at many levels. The degree of detail and the information needed is determined by the decisions to be made.

Teacher A should have information on her performance. This information needs to be in sufficient detail so that she may make a decision as to whether more effort is needed, more training is needed, and what kinds of special efforts are needed to improve performance. However, Principal B may not wish to have quite that much detail even though the decisions made at that level are still decisions about individual teachers. Therefore, individual information in a different format is required.

The project staff will definitely need some information on Teacher A, but summary data on all those teachers who participated in the in-service training program may be far more valuable in making training decisions. As the level or organization broadens, the degree of detail will normally decrease and less and less individualized information will be appropriate.

It is necessary for the evaluator to determine, in advance, what decisions are to be made at what level in order to plan to meet the information requirements of each level. Lack of this advance information on data requirements may result in either a change in the level of decision-making or a change in the kind of decision which is made.

The Indian Education Act mandates evaluation, although it does not specify what kinds of questions should be prepared with respect to in-service training. But questions will be asked, answers will be sought, and the project should be prepared to respond to external as well as internal evaluation needs. A minimal list of such questions might be:

> How many participants attended each in-service training session?
>
> What was the makeup of the session in terms of teachers? Administrators? Indians? Non-Indians?
>
> How much time was devoted to each training session? What materials were used? What were the trainer's qualifications?
>
> What evidence exists that the participants implemented the ideas set forth in the session? To what degree were participants satisfied with the session? Are the effects of the training program noticeable in the classroom?
>
> A carefully designed data collection procedure, a well-documented reporting system, and accurate application of analysis techniques should provide most of the answers.